



Optimización de métodos numéricos en la determinación de la acidificación en los océanos

Antón Velo Lanchas

Julio, 2014

Universidade de Vigo

El Profesor de Investigación Dr. Fiz Fernández Pérez y el Científico Titular Dr. Miguel Gil Coto, del Departamento de Oceanografía del Instituto de Investigaciones Marinas (IIM) perteneciente a la Agencia Estatal del Consejo Superior de Investigaciones Científicas (CSIC), en calidad de directores de tesis, hacen constar:

Que la presente memoria titulada **“Optimización de métodos numéricos en la determinación de la acidificación en los océanos”** presentada por el licenciado Antón Velo Lanchas para optar al grado de Doctor por la *Universidade de Vigo*, fue realizada bajo nuestra dirección y cumple con las condiciones exigidas para su presentación, la cual autorizamos.

Y para que así conste y surta los efectos oportunos, firmamos la presente en Vigo, a 22 de Julio de 2014.



Fdo: Dr. Fiz Fernández Pérez



Fdo: Dr. Miguel Gil Coto

Agradecimientos

A mis directores, Fiz y Miguel, por la ayuda, guía y soporte, pero sobre todo el impulso a esta tesis, ya que no habría sido posible de otro modo.

A Aída y a los coautores de los artículos que componen esta tesis, así como a los proyectos que le han dado soporte.

A mis compañeros en el IIM: Toni, Noe, Merche, Marcos, Paula y Mónica, por toda la ayuda y apoyo.

Y ya en otro plano, Gracias!, como no y sobre todo, a mis padres, amigos y a Mónica por la paciencia y el apoyo incondicional en todo momento.

Índice general

Motivación para la realización de esta tesis	3
Introducción	4
Terminología y acrónimos	5
Capítulo I – Control de calidad de datos	13
1. Descripción de los parámetros	16
1.1. Alcalinidad total	16
1.2. pH	20
2. Metodología	32
3. Umbrales de ajuste	34
4. Evaluación de calidad de datos	35
Capítulo II – Método de interpolación multiparamétrica	45
1. Introducción	45
2. Metodología	49
2.1. Ventana móvil espacial	49
2.2. Distancias multiparamétricas	50
2.3. Algoritmo base	51
3. Evaluación	54
3.1. Oxígeno	60
3.2. Carbono antropogénico.	62
Capítulo III – Estimación de A_T por métodos MLR y neuronales	67
1. Introducción	68

2.	Metodología	70
2.1.	Método 3DwMLR	71
2.2.	Estimación por redes neuronales.	77
3.	Evaluación	81
3.1.	Comparación entre 3DwMLR y red neuronal	83
3.2.	Comparación entre métodos y observaciones	86
Capítulo IV – Estudio de tasas de almacenamiento de C_{ant} en el Atlántico Suroeste		91
1.	Introducción	91
2.	Metodología	92
3.	Evaluación	97
Conclusiones		101
Trabajo actual y futuro		111
Bibliografía		117
Apéndices		127
A.	Capítulo I – CARINA alkalinity data in the Atlantic Ocean	129
B.	Capítulo I – CARINA data synthesis project: pH data scale uni- fication and cruise adjustments	149
C.	Capítulo II - A multiparametric method of interpolation using WOA05 applied to anthropogenic CO_2 in the Atlantic	175
D.	Capítulo III - Total alkalinity estimation using MLR and neural network techniques	189
E.	Capítulo IV – An update of anthropogenic CO_2 storage rates in the western South Atlantic basin and the role of Antarctic Bottom Water	199

Motivación para la realización de esta tesis

La motivación inicial para realizar un doctorado surge tras finalizar la licenciatura, con mi incorporación, bajo la dirección del Dr. Luis Gago Duport al departamento de Xeociencias Mariñas e Ordenación do Territorio de la Universidade de Vigo con la idea de ampliar estudios de cristalización y crecimiento de minerales en fases tempranas de la diagénesis. Dicho estudio finalizó con la obtención del DEA y la presentación de la tesis de licenciatura en el año 2001. En 2007, tras mi incorporación laboral al grupo de CO₂ del departamento de Oceanografía del IIM-CSIC, se abre una nueva posibilidad para desarrollar una tesis doctoral con los Dres. Fiz Fernández Pérez y Miguel Gil Coto, en aspectos técnicos que contribuyan a facilitar la determinación de la acidificación en los océanos, básicamente optimizando e implementando nuevos, o ya existentes, métodos numéricos y rutinas de cálculo claves en este importante proceso marino. El campo es amplio, ya que implica multitud de fases, desde el análisis y procesado de los datos iniciales hasta el procesado final de bases de datos integradas. Así, compaginando y aprovechando tareas a realizar dentro de diversos proyectos en marcha en el grupo, dimos comienzo a la última etapa del doctorado. Este manuscrito es, por tanto, el resultado de dichos trabajos y se presenta como tesis doctoral elaborada con publicaciones. Se trata de artículos que a lo largo de estos años hemos publicado para contribuir, en lo posible, a la mejora del estudio de la acidificación en los océanos, y principalmente como parte del grupo de trabajo internacional CARINA (CARbon in North Atlantic) y en el ámbito de los proyectos CarboOcean y CarboChange.

Introducción

Durante las dos últimas décadas se ha intensificado la observación del océano por diferentes motivaciones científicas auspiciadas por distintos programas nacionales e internacionales. Con ello, se han acumulado una importante cantidad de medidas del sistema del carbónico en agua de mar con diferentes niveles de calidad. El carbono inorgánico total (C_T), la alcalinidad total (A_T) y el pH han sido las variables del sistema del carbónico que con mayor frecuencia se han determinado en la columna de agua. Todas estas medidas necesitan de una evaluación de su calidad, así como de una homogeneización debido a que las medidas de pH se han realizado con escalas y temperaturas diferentes. El objetivo final es conseguir una base de datos homogénea que permita evaluar y 'censar' los cambios en el C_T y pH durante las últimas décadas debido al incremento de CO_2 atmosférico

Dado que para determinar el sistema del carbonato en agua de mar son necesarios únicamente dos parámetros de este sistema, se busca obtener una base de datos general lo más extensa posible que incluya un registro tanto temporal como espacial de la variabilidad del pH y del CO_2 de origen antropogénico. Las bases de datos origen deben ser homogeneizadas y complementadas con la ayuda de diversas metodologías numéricas como paso previo para integrarse en una gran base de datos general.

Es necesario, por lo tanto, elaborar algoritmos que permitan detectar sesgos y valores anómalos en las bases de datos originales. Utilizando diversas técnicas multilineales aplicadas a la variabilidad de parámetros o magnitudes

oceanográficas como temperatura, salinidad, oxígeno y nutrientes, hemos podido discriminar gran parte de la variabilidad real de la generada por los sesgos producidos por errores en las metodologías analíticas. Finalmente, desarrollando nuevos algoritmos de interpolación se ha propagado de forma optimizada la información disponible, originalmente sin estructura geográfica homogénea, a una red geográfica regular para generar un producto que facilite la evaluación regional del almacenamiento de CO₂ y de la variabilidad del pH en el océano.

Terminología y acrónimos

A continuación se presenta un resumen de terminología empleada para facilitar la lectura.

Términos genéricos

Parámetro: También se emplea la palabra Variable. Refiere a cualquiera de un conjunto de propiedades físicas, químicas o de otro tipo, cuyos valores determinan las características o el comportamiento de un sistema particular. Ejemplo: Alcalinidad, Carbono Inorgánico Total, pH, etc.

Outlier: Valor anómalo o atípico para un determinado parámetro. Un valor puede identificarse como anómalo cuando está fuera del rango posible para ese parámetro (ej. agua líquida a $T > 100^{\circ}\text{C}$ y presión atmosférica) o cuando se puede considerar que está fuera del rango esperable para ese punto.

Términos relacionados con bases de datos

CARINA: (CARbon in North Atlantic) Proyecto de síntesis de datos de carbono y trazadores financiado por el EU IP CARBOOCEAN. Comparte nombre con las bases de datos que son objetivo del proyecto, y consiste en una síntesis de datos de botella de 188 campañas oceanográficas, relevantes al sistema del carbono. Su principal área de cobertura es el Océano Atlántico y Ártico, y presenta datos hasta el 2005.

ETOPO2v2: Conjunto de datos de relieve (topografía/batimetría) de la agencia NOAA. Presentan una cobertura global y una resolución de 2 minutos de grado. Se emplean como referencia para el fondo oceánico.

GLODAP: (GLObal Ocean Data Analysis Project) Esfuerzo cooperativo financiado por NOAA, DOE, NSF y JGOFS-SMP con el objetivo de realizar una síntesis global de datos de carbono y trazadores. Emplea principalmente datos de campañas de WOCE, JGOFS y OACES/NOAA sobre la década de 1990.

WOA: (World Ocean Atlas) Conjunto de datos publicados por NODC/NOAA que incorpora temperatura, salinidad, oxígeno disuelto, utilización aparente de oxígeno (AOU), porcentaje de saturación de oxígeno, fosfato, silicato y nitrato. Los datos representan una climatología realizada mediante análisis objetivo sobre los datos de partida del WOD (World Ocean Database, NODC/NOAA) y se ofrecen en una cuadrícula global de 1° (Latitud/Longitud) y a profundidades estándar. Ofrece productos anuales, estacionales y mensuales.

Términos relacionados con control de calidad

QC1: También denominado 1stQC o control de calidad primario. Es la primera fase de control de calidad y se realiza individualmente sobre cada campaña. El análisis intenta detectar datos anómalos y errores obvios como errores de escala, de transcripción, etc. Intenta incluso verificar en lo posible la consistencia interna de los parámetros disponibles para la campaña. Se trata, por tanto, de una revisión de la precisión (Inglés: *precision*) de cada campaña individual.

QC2: También denominado 2ndQC o control de calidad secundario. Es la segunda fase de control de calidad y se realiza secuencialmente tras la primera. Los análisis pretenden identificar diferencias sistemáticas entre campañas que no sea posible asimilar con la variabilidad natural a lo largo del tiempo. Se trata, por tanto, de un análisis de consistencia entre campañas que pretende revisar la exactitud (Inglés: *accuracy*) de cada campaña.

Crossover: Comparativa de cualquier parámetro entre un par de campañas oceanográficas (en adelante denominadas A y B) próximas, o cuyo transecto(s) se cruza(n). La comparativa se realiza entre perfiles de estación.

Offset: El resultado numérico del crossover. Representa la cantidad por la cual los datos de un parámetro de la campaña B está desplazada (offset) con respecto a los de la campaña A. Los métodos de evaluación del Crossover también generan un estadístico de la incertidumbre para ese Offset.

Inversión: La aplicación de una metodología para, a partir del conjunto de offsets y sus correspondientes incertidumbres, calcular las correcciones óptimas que se deberían aplicar a campañas individuales para minimizar los desajustes entre el grupo de campañas estudiado. Se emplean dos procedimientos, uno denominado mínimos cuadrados ponderados (*Weighted Least Squares*: WLSQ) que toma las incertidumbres como factor de ponderación en el ajuste de mínimos cuadrados, y otro, denominado mínimos cuadrados ponderados amortiguados *Weighted Damped Least Squares*: WDLSQ).

Corrección: El valor resultado de la(s) inversión(es) para cada campaña, es decir, el valor de la corrección que la(s) inversión(es) propone(n) para cada campaña.

Ajuste mínimo: No todas las correcciones que generan las inversiones se aplican. Entre otros criterios, se define un umbral mínimo para cada parámetro. Por debajo de este límite no se considera necesario aplicar las correcciones y se conservan los datos originales.

Ajustes: Valores finales de corrección aplicados a los parámetros de cada campaña. La corrección generada por las inversiones se toma como referencia pero se realiza una supervisión manual final y es necesario un acuerdo dentro del grupo CARINA para cada ajuste individual.

Flag: Etiqueta o indicador de calidad de un dato. Generalmente es un valor numérico entero. Valores comunes en el contexto del proyecto CARINA son: 0, que indica que el dato es calculado o interpolado; 2, que indi-

ca dato aceptable o bueno; 3, que indica dato cuestionable o 4 malo; y 9 que indica que no hay dato. Una referencia completa para flags usados en CARINA se puede encontrar en Tanhua et al. (2010b). Unas recomendaciones generales se pueden encontrar en el documento Ocean Data Standards, Vol.3 (Intergovernmental Oceanographic Commission of UNESCO, 2013).

Términos relacionados con metodologías

MLR: Regresión multilíneal. Modelo lineal que describe la relación entre una única variable denominada dependiente o respuesta y un conjunto de variables denominadas independientes, explicativas o predictoras. El concepto “multi” deriva de la presencia de múltiples variables independientes, y el concepto “lineal” de que la variable respuesta es una función lineal de los coeficientes aplicados a las variables independientes. El cálculo de los coeficientes se realiza a partir de una serie de observaciones de las variables independientes y dependiente que se someten a un procedimiento de mínimos cuadrados según una relación multilíneal propuesta. Una descripción más detallada se puede encontrar en Kutner et al. (2004).

Robust regression: Método de análisis de regresión similar al MLR y donde el concepto “robusto” deriva del empleo de técnicas para la detección y el ponderado de valores anómalos. Hay diversas funciones para realizar el ponderado, pero como norma general, con todas ellas se aplica un procedimiento iterativo que analiza la contribución de cada valor indi-

vidual y reitera modificando el peso de cada valor hasta que el ajuste no mejora significativamente. Suele reemplazar estadísticos clásicos por otros más robustos a valores anómalos como, por ejemplo, la mediana en lugar de la media.

Redes Neuronales: Las redes neuronales se pueden definir, según Sarle (1994), como una amplia clase de flexibles regresiones no lineales, modelos discriminantes, modelos de reducción de datos, y sistemas dinámicos no lineales. Consisten, a menudo, en un número elevado de “neuronas”, es decir, elementos de computación simples, lineales o no lineales, interconectados de modos habitualmente complejos y, habitualmente también, organizados en capas.

Capítulo I

Capítulo I – Control de calidad de datos

Los trabajos que componen esta tesis, se engloban principalmente en el marco del proyecto CARINA (CARbon in North Atlantic), que nació de modo informal y sin financiación en Kiel (Alemania) en 1999, con el objetivo de recopilar y generar una base de datos para el Océano Atlántico Norte de parámetros relacionados con el carbono que permitiera calcular de manera precisa los inventarios y tasas de captación de carbono por parte de los océanos.

La base de datos generada recibe el mismo nombre que el proyecto: CARINA. Es, por tanto, una base de datos del sistema del carbonato y de parámetros auxiliares que se necesitan para estudiar la variabilidad del CO₂ en agua de mar, generada a partir de datos hidrográficos de campañas oceanográficas realizadas en los océanos Atlántico, Ártico y Austral, aunque incluye también algunas campañas más dispersas realizadas en el Pacífico. Además de la propia recolección de datos, el control de calidad ha sido uno de los objetivos principales del proyecto, realizándose un control de calidad primario (QC1) y secundario de los datos (QC2). Los resultados del proyecto CARINA se pueden dividir en dos bloques:

El primer bloque consiste en una base de datos de ficheros individuales de cada campaña, donde se almacenan todos los datos medidos y sus etiquetas

(*flags*) de calidad. Estos ficheros se generaron en formato de intercambio WHP (*WOCE Hydrographic Program*), que incluye los metadatos condensados en las primeras líneas. En esencia no se incluyen datos calculados ni interpolados en estos ficheros, con las excepciones de la profundidad (estimada a partir de presión y latitud), y la salinidad (obtenida con conductividad, temperatura y presión). Tampoco se han aplicado correcciones o ajustes a ninguno de los parámetros almacenados en estos ficheros, con la excepción de las medidas de pH, que sí se han unificado en una única escala, la denominada “*Seawater*” (SWS) a 25°C.

El segundo bloque de CARINA es el producto de síntesis como tal del proyecto. Son tres ficheros de datos de campañas geográficamente agrupados, con control de calidad realizado y ajustes aplicados. Las agrupaciones son Océano Atlántico (denominado ATL), Océano Ártico y Mares Mediterráneos (denominado AMS) y Océano Austral (SO). Estos ficheros incluyen:

1. Valores medidos provenientes de los ficheros individuales de cada campaña, con el control de calidad primario (QC1) y el secundario (QC2, ajustes) realizado.
2. Valores interpolados para nutrientes, oxígeno y salinidad en aquellos registros que no contengan dichos datos y siempre que sea posible realizar la interpolación según los criterios que se describen en Key et al. (2010).
3. Parámetros de carbono calculados mediante las ecuaciones del sistema del carbonato donde sea posible.

Los valores calculados (3) o interpolados (2) que se incluyen en el producto se marcan con la etiqueta de calidad “0”, para así identificarlos.

Todos los datos incluidos en el producto de síntesis han sido corregidos según los criterios establecidos en el proyecto CARINA, es decir, tras revisión y acuerdo de sus integrantes. Incorporan, por lo tanto, el ajuste acordado para cada campaña. En los correspondientes artículos (Velo et al., 2010a, 2009) se detallan y fundamentan, para los diferentes parámetros o agrupaciones de estos, cada uno de los ajustes propuestos para las campañas en las que, mediante los controles de calidad, se detectaron sesgos. En muchos casos, en los ficheros individuales de partida, se dispone de valores de otros parámetros (p. ej. ^{14}C , ^{13}C y SF_6) que no han sido incluidos en el control de calidad secundario.

En los dos primeros artículos de esta tesis (Velo et al., 2010a, 2009), se describen los análisis de consistencia de las medidas de alcalinidad y pH de la región del Océano Atlántico de la base de datos de CARINA (CARINA-ATL). Para el caso del pH el trabajo se realizó a nivel global para los tres productos de datos, y fue necesario realizar además y previamente, una homogenización o unificación de escalas.

Una descripción más detallada del proyecto CARINA y sus bases de datos se puede encontrar en Key et al. (Key et al., 2010)* y Tanhua et al. (Tanhua et al., 2009)*.

Una visión de conjunto del Atlántico Norte se incluye en: Tanhua et al. (Tanhua et al., 2010a)*.

Información sobre el control de calidad en alcalinidad para otras áreas se puede encontrar en:

- **Océano Ártico** en Jutterström et al. (2010)
- **Mares Nórdicos** en Olsen et al. (2009)
- **Sector Atlántico del Océano Antártico** en Hoppema et al. (2009)*
- **Sector Índico del Océano Antártico** en Lo Monaco et al. (2010)*
- **Sector Pacífico del Océano Antártico** en Sabine et al. (2009)

Los métodos y técnicas de control de calidad están descritos en detalle en el artículo de métodos del especial de ESSD (Tanhua et al., 2010b)* y en el paquete de herramientas de Matlab (Tanhua, 2010) que implementa dichas técnicas y métodos.

Se indican con un asterisco (*) los artículos participados por el autor y que contribuyen a explicar o continúan las tareas de este capítulo.

A continuación se describen en detalle las características de los parámetros analizados

1. Descripción de los parámetros

1.1. Alcalinidad total

La alcalinidad total (A_T , TALK) en agua de mar se define como "...el número de moles de ion hidrógeno equivalente al exceso de aceptores de protones (bases formadas de ácidos débiles con constantes de disociación K

$\leq 10^{-4.5}$ a 25°C y cero fuerza iónica) sobre donantes de protones (ácidos con $K > 10^{-4.5}$) en 1 kg de muestra.” (Dickson, 1981). La alcalinidad total es uno de los cuatro parámetros básicos del sistema del dióxido de carbono en agua de mar, siendo los otros tres: el carbono inorgánico total disuelto (C_T , DIC, TIC, TCO_2 , $TCARBN$), la fugacidad del CO_2 disuelto (fCO_2) y el pH. Si al menos dos de estos parámetros son conocidos, los parámetros restantes pueden ser calculados usando constantes termodinámicas para una temperatura, salinidad y presión determinadas.

Disponer de datos de carbono de alta calidad es crítico para poder detectar cambios pequeños del sistema del CO_2 . Concretamente, dos parámetros, la alcalinidad total y el carbono inorgánico total, se emplean en bastantes metodologías basadas en técnicas de retro-cálculo para estimar la señal del CO_2 antropogénico. La alcalinidad total es clave tanto para determinar los cambios en carbono total producidos por la disolución de carbonato cálcico (Feely et al., 2002) como para establecer las concentraciones de carbono total en aguas superficiales en equilibrio con la atmósfera, y tanto para niveles de CO_2 históricos como presentes o futuros. Es necesario resaltar también que muchas medidas del sistema del CO_2 en agua de mar fueron realizadas mediante pares A_T -pH para determinar C_T , y, por esta razón, ambos parámetros son claves para los objetivos de CARINA.

Los datos de A_T incluidos en el conjunto de datos de CARINA-ATL tienen su origen en múltiples grupos internacionales de investigación que emplean varios métodos de análisis diferentes. La mayoría de datos de A_T fueron determinados empleando un método de valoración potenciométrica en celda

cerrada (Dickson et al., 2007) y otra gran parte se obtuvo mediante métodos de valoración potenciométrica a punto final en celda abierta (Mintrop et al., 2000).

Distribución

La distribución espacial de la alcalinidad en el océano, viene determinada principalmente por la variabilidad de dos parámetros, que son la salinidad y el silicato. En las capas superficiales del océano, la relación entre la alcalinidad y salinidad ha sido ampliamente tratada en la literatura (Lee et al., 2006; Millero et al., 1998), y se debe principalmente a los procesos de dilución y concentración del agua de mar que se producen por la adición de agua dulce a partir de precipitaciones y fusión de hielo, y por la retirada de agua debida a la evaporación o la formación de hielo. En los paneles (a) y (b) de la Figura 1.1 se muestra la alcalinidad y la salinidad para una sección norte-sur del Océano Atlántico. En general la correlación positiva entre ambas es notoria y ligada al ciclo del agua (evaporación en los giros subtropicales y precipitación neta hacia las zonas polares).

En la columna de agua, la síntesis y remineralización de materia orgánica añade o elimina iones (nitrato fundamentalmente) de la disolución, afectando a la alcalinidad (Brewer et al., 1975; Fraga and Álvarez-Salgado, 2005), aunque de manera minoritaria, a lo largo de los primeros 1000-1500 metros de la columna de agua, en los cuales este proceso es dominante frente al de disolución de CaCO_3 . Si analizamos la alcalinidad normalizada (NTA) (Panel c de la Figura 1.1), que excluye el efecto de la sal (Millero et al., 1998),

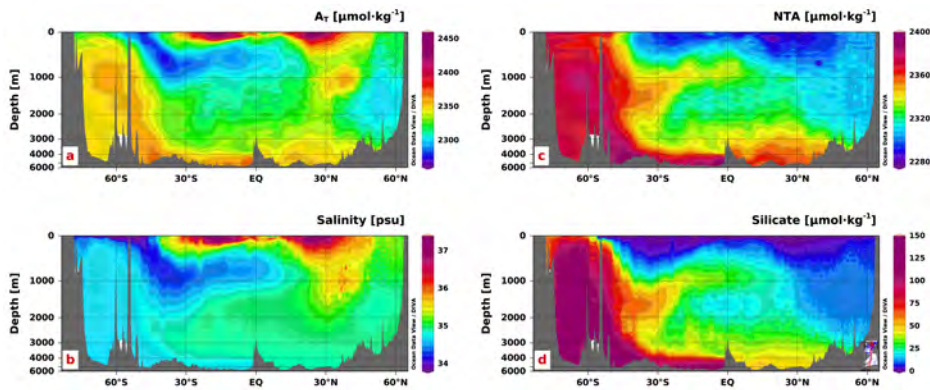


Figura 1.1: Sección Norte-Sur del Océano Atlántico donde se muestra la alcalinidad total (panel a), salinidad (panel b), alcalinidad normalizada (panel c) y silicato (panel d). La escala vertical de profundidad no es lineal para exagerar la representatividad de las capas más someras.

se puede observar que ésta presenta una correlación muy alta con el silicato (Panel d de la Figura 1.1) en las aguas asociadas a la AABW, lo que denota que la circulación profunda está afectando a ambos parámetros. La disolución del ópalo (sílice) generado por diatomeas presentes en gran cantidad en las aguas australes motiva el incremento de silicato en estas aguas. Asimismo, la acumulación de la disolución de organismos calcáreos tanto en el Pacífico como en la Antártida provoca también un aumento de la alcalinidad en las aguas presentes en la Antártida. A partir de ahí la circulación oceánica global traslada hacia el norte, desde la Antártida al Océano Atlántico, los altos valores de silicato y NTA. Se observan también en aguas profundas valores altos de NTA sin el consiguiente aumento de silicato, ello es debido a la disolución de CaCO_3 en las zonas profundas del Atlántico tropical. La producción biogénica de carbonato en aguas superficiales y el hundimiento de estos esqueletos y estructuras formadas, provoca la llegada de carbonato

en diferentes formas minerales a estas aguas profundas donde lentamente es disuelto (Wolf-Gladrow et al., 2007). En algunos casos es difícil distinguir si este proceso ocurre realmente en la columna de agua, o si es debido a la interacción con el sedimento en las masas de agua próximas a él (Berelson et al., 2007).

Algunos autores han introducido además la temperatura del agua para la parametrización de la alcalinidad, aunque la fracción de variabilidad explicada es mucho más pequeña que la determinada por la salinidad (Millero et al., 1998; Vázquez-Rodríguez et al., 2012); en esa relación subyace la covariación entre, por un lado, aguas tropicales y subtropicales en las cuales los organismos calcáreos substraen iones carbonato para formar su estructuras y, por otro lado, aguas frías de altas latitudes donde estos organismos son minoritarios.

1.2. pH

Los principales factores que gobiernan la distribución espacial y temporal del pH oceánico son la temperatura, debido a la dependencia de las constantes de disociación con este parámetro, junto con la presión parcial de CO_2 ($p\text{CO}_2$), la alcalinidad total (A_T) y el carbono total (C_T), estos últimos debido a su participación en el equilibrio del carbonato. El pH oceánico superficial se ve afectado principalmente por la temperatura, la biología, el intercambio de CO_2 en ambas direcciones con la atmósfera y el intercambio de CO_2 con aguas más profundas debido a procesos de afloramiento y hundimiento. Hay cambios y ciclos que contribuyen a las fluctuaciones de pH en la superficie del océano, en escalas de tiempo que van de días a años, incluyendo el ciclo estacional

(Raven et al., 2005; Wootton et al., 2008). Se estima además que las aguas superficiales del océano se han acidificado, globalmente, en un $\approx 34\%$, es decir, unas 0.13 unidades de pH (de 8.2 a 8.1 en escala SWS a 25°C) desde los tiempos pre-industriales (Caldeira and Wickett, 2003; Olafsson et al., 2009; Orr et al., 2005; Raven et al., 2005), debido a emisiones antropogénicas de CO_2 y su captación por el océano. En el océano profundo la concentración de CO_2 se incrementa por la abundancia de CaCO_3 en sedimentos y su disolución en la columna de agua, aumentando el efecto tampón y manteniendo estable el pH del océano profundo, incluso conduciendo a cambios mínimos a escalas de tiempo de 10000 años (Raven et al., 2005).

En la Figura 1.2 se muestra la distribución de cuatro parámetros para una sección norte-sur en el Océano Atlántico. Se ha exagerado la escala vertical para mostrar mejor los efectos en las capas menos profundas, y se ha utilizado una escala inversa de colores para pH para mostrar mejor la concordancia. En primer lugar se puede observar en los paneles (a) y (b) que la mayor parte de la variabilidad del pH está ligada al consumo y producción de nitrato asociado con la actividad fotosintética en superficie y la remineralización en la columna de agua, fundamentalmente en los primeros 1500 metros.

En el panel (c) se observa la alta correlación del pH con la relación C_T/A_T . Se aprecia también como el efecto tampón del buffer de carbonato provoca que el pH apenas se vea afectado por la disolución de CaCO_3 en las capas más profundas, efecto que sí se puede apreciar en el C_T (panel d) o la A_T y NTA (Figura 1.1, paneles a y c) por separado. Esto es debido a que por cada lado la variabilidad del pH y la A_T correlacionan independientemente con los dos

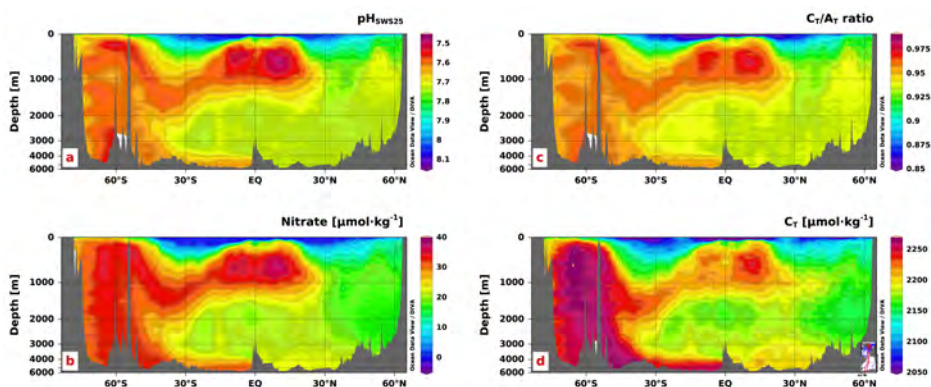


Figura 1.2: Sección Norte-Sur del Océano Atlántico donde se muestra el pH en escala SWS a 25°C (panel a), nitrato (panel b), relación C_T/A_T (panel c) y carbono total (panel d). La escala vertical de profundidad no es lineal para exagerar la representatividad de las capas más someras.

procesos básicos de la bomba biológica. Por un lado la remineralización de la materia orgánica, que afecta al pH pero muy poco a la A_T y por otro lado la disolución de CaCO_3 que afecta muy poco al pH pero muy marcadamente a la A_T . La combinación de ambos presenta una ventaja notable para determinar la variabilidad espacial del C_T .

En cuanto al efecto de la disolución de CO_2 atmosférico y la temperatura sobre el sistema del carbonato en general y sobre el pH del agua de mar en particular, se incluyen a continuación dos Figuras (1.3 y 1.4) que ilustran la respuesta de este sistema ante los mencionados cambios.

La Figura 1.3 muestra en el panel superior la clásica evolución de la concentración de CO_2 en la atmósfera para el observatorio de Mauna Loa (Hawaii) (Tans and Keeling, 2014). En color verde se presenta un promedio mensual, que muestra las variaciones del ciclo estacional, y en azul un promedio anual, donde se puede observar la evolución de la $f\text{CO}_2$ atmosférica desde 1960 hasta

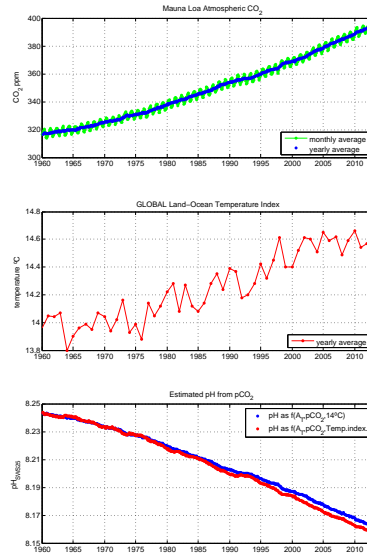


Figura 1.3: Panel superior: evolución de fCO₂ observada en el observatorio de Mauna Loa (Hawai) para el período 1960-2013. Panel medio: Evolución de temperatura a partir del Índice de Temperatura Tierra Océano (NASA/GISS, 2014). Panel inferior: Estimación mediante ecuaciones termodinámicas (van Heuven et al., 2009) del cambio de pH para el cambio de condiciones descrito en los paneles superiores (se considera sin variación al resto de parámetros)

2013, siempre ascendente y de tendencia exponencial (Conocida como curva de Keeling). En el segundo panel se muestra la evolución de la temperatura promedio global para cada año obtenida como anomalía a partir del Índice de Temperatura Tierra-Océano (LOTI), que combina mediciones terrestres y oceánicas (NASA/GISS, 2014). Se toma como base una temperatura de 14°C (NASA/GISS, 2014). En el panel inferior se muestra una estimación de la evolución del pH del agua de mar modelado (van Heuven et al., 2009) como función de un valor constante de A_T para el intervalo 1960-2013 (2350 μmol·kg⁻¹). La línea azul muestra la evolución del pH considerando única-

mente la diferencia de $f\text{CO}_2$ y empleando una temperatura constante (14°C). La línea roja tiene en cuenta, además de la evolución de $f\text{CO}_2$, el incremento de temperatura observado a lo largo del período. El panel inferior simularía, por tanto, la evolución del pH en aguas superficiales en equilibrio de CO_2 con la atmósfera, y considerando únicamente las variaciones en $f\text{CO}_2$ o [$f\text{CO}_2$ y temperatura] registradas. Se observa como el incremento de $f\text{CO}_2$ en la atmósfera provoca, por sí solo, un descenso del pH de 81 milésimas a lo largo del intervalo 1960-2013. El incremento promedio de temperatura provocaría un descenso adicional de 5 milésimas, dando un total de 86 milésimas de descenso de pH en base únicamente a cálculos de equilibrio termodinámico.

La Figura 1.4 muestra una simulación del sistema del carbonato a partir de valores medidos de A_T para la capa de 100 m y valores del producto en cuadrícula de WOA09 para temperatura, salinidad y nutrientes (Antonov et al., 2010; Garcia et al., 2010; Locarnini et al., 2010). En los paneles (a) y (b) se muestran los datos en cuadrícula de la climatología anual de WOA09 correspondientes a la capa de 100 m para temperatura (panel a) y salinidad (panel b). En el panel (c) se muestran datos de A_T para la capa de 100 m el Océano Atlántico. Se partió de datos medidos de campañas oceanográficas de los proyectos CARINA, GLODAP, CLIVAR y otras campañas realizadas por el IIM-CSIC, además de datos estimados por MLR (Capítulo III: Velo et al., 2013) que se interpolaron mediante el método WMP (Capítulo II: Velo et al., 2010b). En los paneles (d) y (e) se presentan dos mapas con modelizaciones de valores de pH para la capa antes indicada del océano, partiendo de los datos termohalinos y bioquímicos de WOA09, los datos de

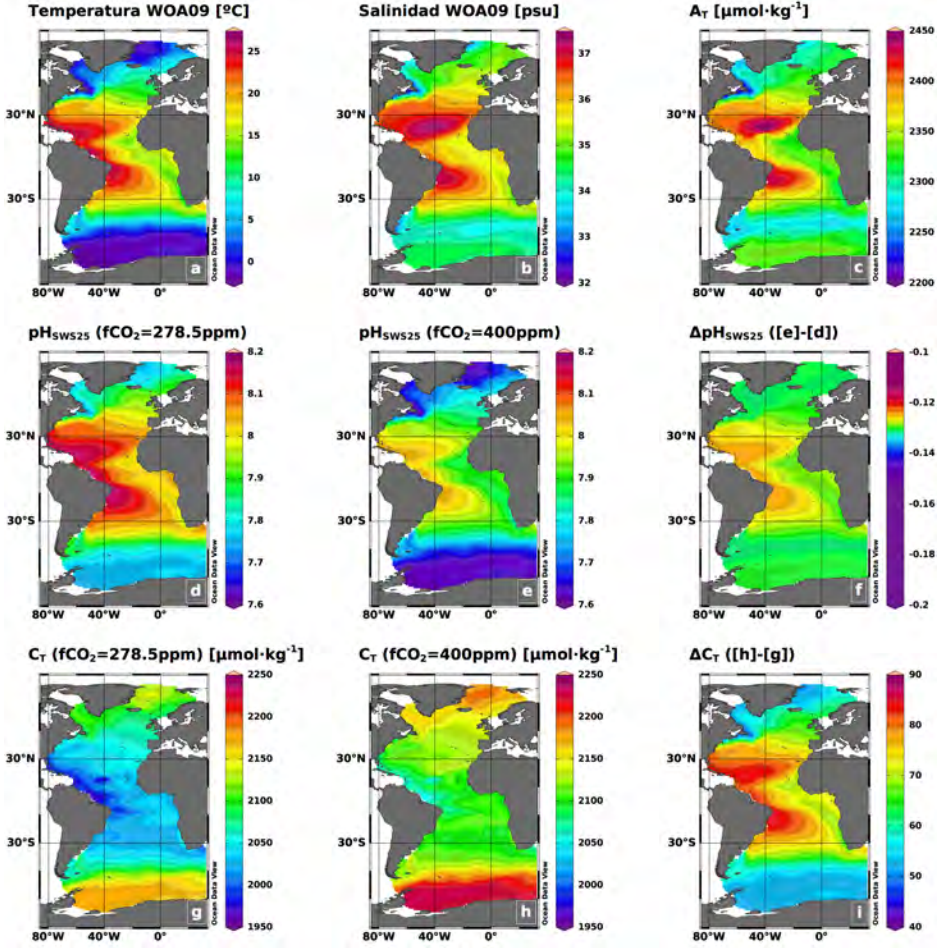


Figura 1.4: Simulación de pH en condiciones preindustriales ($f\text{CO}_2=278.5$ ppm) y actuales ($f\text{CO}_2$ 2014 ≈ 400 ppm). Paneles (a y b): Datos extraídos de la capa de 100 m de la climatología de WOA09 para temperatura (a) y salinidad (b). Panel (c): Datos extraídos de la capa de 100 m generada mediante interpolación WMP de datos de botella de CARINA+GLODAP+CLIVAR+Campanías IIM-CSIC. Paneles (d, e, f): Simulación de pH en la capa de 100 m para condiciones en equilibrio con $f\text{CO}_2$ de 278.5 ppm (d), 400 ppm (e) y diferencia entre ambas (f). Paneles (g, h, i): Simulación de C_T en la capa de 100 m para condiciones en equilibrio con $f\text{CO}_2$ de 278.5 ppm (g), 400 ppm (h) y diferencia entre ambas (i).

A_T del panel (c), y una $f\text{CO}_2$ constante de 278.5 ppm (preindustrial) para el panel (d) y 400 ppm (aproximadamente equivalente a 2014) para el panel

(e). En el panel (f) se muestra la diferencia de pH obtenida para ambas simulaciones (400 ppm-278.5 ppm). En los paneles (g) y (h) se presentan los dos mapas equivalentes a los paneles (d) y (e) con los valores de C_T correspondientes a esas mismas modelizaciones. En el panel (i), equivalente al (f) se muestra la diferencia de C_T obtenida para ambas simulaciones (400 ppm-278.5 ppm). En los paneles en los que se muestran las simulaciones se ha intentado preservar una concordancia en la escala. De este modo, la escala de colores para los paneles (d) y (e) es la misma en ambos (7.6-8.2), al igual que ocurre con (g) y (h) ($1950-2250 \mu\text{mol}\cdot\text{kg}^{-1}$) permitiendo así apreciar la magnitud de los cambios entre escenarios. Los paneles con las diferencias (f) e (i) presentan una escala concordante con los rangos anteriores pero notablemente más reducida; en concreto se ha reducido a la sexta parte del rango indicado anteriormente para los parámetros, resultando en (-0.1 a -0.2) para pH y ($40-90 \mu\text{mol}\cdot\text{kg}^{-1}$) para C_T . Se puede observar que pese a esta concordancia ha sido necesario exagerar la escala de colores para pH, haciéndola no lineal alrededor del valor mediana obtenido (-0.130)

Analizando la figura en detalle, se puede observar, en primer lugar, la fuerte dependencia del pH con la temperatura para ambos escenarios: el equivalente a 278.5 ppm (panel d) y el de 400 ppm (panel e). Analizando los valores de correlación, obtenemos un ajuste estadístico con un R^2 prácticamente de 1 (0.9998) y con unas pendientes muy similares (0.0158 y 0.0161 unidades de pH/°C respectivamente) para ambos casos.

En segundo lugar, analizando el panel (f) que ilustra el cambio de pH entre ambos escenarios, se puede ver como el incremento de $p\text{CO}_2$ afecta

de un modo prácticamente constante a la acidificación, siendo la variación de -0.131 ± 0.001 unidades de pH para temperaturas inferiores a 10°C y de entre -0.130 y -0.124 unidades de pH para temperaturas entre 10°C y 25°C (-0.129 ± 0.04 unidades de pH para el conjunto de datos).

En tercer lugar, comparando los paneles (d y e) que muestran los escenarios de pH frente a los de C_T (g y h) se puede observar como existe una correlación inversa entre ambos parámetros, con cambios muy similares pero influenciados por la señal de la A_T . En ausencia de otros factores de cambio, la alcalinidad y el carbono total correlacionan positivamente entre sí, y esta correlación es, sin embargo, negativa frente al pH, por lo que los efectos se compensan parcialmente. Como ya se mostró en los paneles (a) y (c) de la Figura 1.2, el factor C_T/A_T , que normaliza el carbono total frente a la alcalinidad, ilustra la correlación inversa del factor con el pH debida a esa compensación. En el caso de los paneles (g y h) de C_T se observa como el efecto de compensación provoca la desaparición de los valores máximos de alcalinidad y pH en zonas tropicales. Aun así, se puede apreciar en cierto dominio del pH.

Por último, analizando los paneles (g) e (i) que respectivamente muestran los cambios de pH y C_T motivados por el cambio de $f\text{CO}_2$, se puede observar que el cambio de C_T correlaciona positivamente con la temperatura (T), con un R^2 de 0.994 y un factor $\Delta\Delta C_T/\Delta T$ de $1.24 \mu\text{mol}\cdot\text{kg}^{-1}\cdot^\circ\text{C}^{-1}$. Esto es debido a la variación del efecto tampón del sistema del carbonato, que es menor en aguas más frías. Analizando las zonas tropicales, con incrementos de C_T próximos a $80 \mu\text{mol}\cdot\text{kg}^{-1}$ se observa que los cambios producidos en pH (≈ -0.124 unidades de pH) son muy próximos a los producidos en zonas

polares (≈ -0.130 unidades de pH) con incrementos en C_T mucho menores ($\approx 50 \mu\text{mol}\cdot\text{kg}^{-1}$). Los cambios en el ΔpH están influenciados por la variabilidad espacial de la A_T (notoria en latitudes al sur del paralelo 30°S) y son 2 órdenes de magnitud inferiores a la variabilidad natural de la distribución del pH; esta variación es debida a la no linealidad del sistema del carbonato.

Escalas de pH

La primera definición de pH de Sørensen (1909) presenta varios problemas operacionales debido a que los protones libres $[\text{H}^+]$ no existen en cantidades significativas en soluciones acuosas. Por ello, el símbolo " H^+ " representa, en realidad, hidratos complejos en lugar de la concentración de iones hidrógeno libres. La primera definición operacional fue la denominada escala NBS (Bates and Vijh, 1973) acrónimo del *National Bureau of Standards*. La escala de pH_{NBS} se define por unas series de soluciones tamponadas estándar con valores de pH asignados cercanos a "*the best estimates of the proton activity (a_{H^+})*," de modo que el $\text{pH}_{\text{NBS}} = -\log(a_{\text{H}^+})$. El estado de referencia para la escala NBS es la solución infinitamente diluida, lo cual es muy útil en aguas naturales muy diluidas, como las de ríos y lagos. Sin embargo, esta escala no es recomendada para agua de mar, debido a su gran fuerza iónica (Dickson, 1984; Millero et al., 1993).

Además de la escala NBS, se han sugerido otras tres escalas para agua de mar. La escala "*Free hydrogen ion*" (pH_F), la escala "*Total hydrogen ion*" (pH_T) y la escala "*Seawater*" (pH_{SWS}). La razón de la existencia simultánea de cuatro escalas de pH es principalmente histórica, y refleja el refinamiento

gradual de la determinación experimental de pH en agua de mar. Las definiciones de las diferentes escalas se resumen a continuación en la Tabla I-1, reproducida de (Velo et al., 2010a).

La escala “Free” es conceptualmente la más clara, definida únicamente por la concentración de H^+ . La contrapartida es que el H^+ no puede ser medido directamente. Esta inconveniencia operacional se resuelve con las escalas *Total* y *Seawater*.

Escala de pH	Definición
NBS	$pH_{NBS} = -\log(a_H)$
Free	$pH_F = -\log[H^+]_F$
Total	$pH_T = -\log[H^+]_T = -\log([H^+]_F \cdot (1 + [SO_4^{-2}]/K_S))$ $\approx -\log([H^+]_F + [HSO_4^-])$
Seawater	$pH_{SWS} = -\log[H^+]_{SWS} = -\log([H^+]_F \cdot (1 + [SO_4^{-2}]/K_S + [F^-]/K_F))$ $= -\log([H^+]_F + [HSO_4^-] + [HF])$

Tabla I-1: Escalas de pH empleadas en mediciones en agua de mar, definiciones y relaciones (Dickson, 1993; Dickson et al., 2007; Millero, 2007)

La escala *Total* de pH (Hansson, 1973) tiene en cuenta la asociación del ion SO_4^- con H^+ para formar HSO_4^- (incluyendo el SO_4^{-2} en sus soluciones de calibración), evitando la definición de la constante de disociación del HSO_4^- cuyo valor es difícil de obtener en agua de mar. Por otra parte, la escala *Seawater* (Dickson and Riley, 1979) incluye, además del bisulfato, la reacción de asociación del ion fluoruro.

Por lo tanto, las diferencias entre las escalas *Total* y *Seawater* derivan de la inclusión o no del fluorhídrico en los medios en los cuales se basa la escala. Sin embargo, esta última diferencia es pequeña (sobre $\approx 0.0095 \pm 0.0005$ unidades de pH a salinidad 35) debido a que la concentración de HSO_4^- es mucho mayor que la de HF en agua de mar. Al contrario, el pH reportado en escala “Free” es

sobre 0.11-0.12 unidades de pH más alto que en las escalas *Total* o *Seawater* (Zeebe and Wolf-Gladrow, 2001). Debemos recordar que estas diferencias son mucho mayores que la precisión actual alcanzada en las mediciones de pH, que son del orden de ± 0.0004 -0.001 unidades de pH (Clayton and Byrne, 1993) y son, por lo tanto, significativas.

Los datos de pH incluidos en CARINA provienen de la recopilación de datos de campañas de múltiples grupos internacionales hasta el año 2005. Los datos incluyen medidas obtenidas empleando métodos potenciométricos y espectrofotométricos, y han sido reportados en distintas escalas de pH y varias combinaciones de temperatura y presión. En total, la base de datos de CARINA tiene datos de pH de 3761 estaciones de 59 campañas, resultando en 49915 medidas de pH. Se puede encontrar una referencia completa de las campañas incluidas así como las escalas y condiciones empleadas en la Figura 3 del artículo Velo et al. (2010a).

Conversión a pH SWS 25°C

Como primer paso para el control de calidad entre campañas, todos los valores de pH fueron unificados a la escala *Seawater* a 25°C. La elección de esta escala se efectuó por convención en reuniones del grupo de CARINA. La conversión de escala se efectuó mediante las rutinas de software CO2SYS (Lewis et al., 1998) convertidas a paquete MATLAB (van Heuven et al., 2009). Este paquete, que se puede obtener en <http://cdiac.ornl.gov/oceans/co2rprt.html>, implementa las ecuaciones del carbonato en agua de mar y realiza todos los cálculos necesarios para resolver este sistema. Así,

Ecuaciones empleadas en la conversión de escalas	
$pH_T = pH_{SWS} - \log((1 + S_T/K_S)/(1 + S_T/K_S + F_T/K_F))$	
$pH_T = pH_F - \log(1 + S_T/K_S)$	
$pH_T = pH_{NBS} - (\log(1 + S_T/K_S) + \log(f_H))$	

Tabla 1-2: Ecuaciones empleadas por el paquete Matlab CO2SYS para la conversión de escalas de pH. Las rutinas emplean internamente la escala *Total* para los cálculos del sistema del carbonato.

esta rutina de código permite calcular el pH en cualquiera de las cuatro escalas empleadas en agua de mar y, asimismo, permite al usuario escoger el paquete de constantes que desea emplear para los sistemas del carbonato y borato.

Los parámetros adicionales que se requieren para estos cálculos son la presión, la salinidad, el silicato, el fosfato y la alcalinidad. Para la conversión se emplearon las siguientes constantes: las constantes del ácido carbónico en agua de mar y escala NBS de Mehrbach reajustadas por Dickson y Millero (Dickson and Millero, 1987; Mehrbach et al., 1973) a la escala SWS y las del bórico y bisulfato de Dickson (Dickson, 1990a, 1990b). Internamente, la rutina de MATLAB convierte inicialmente todos los valores de pH a la escala *Total* utilizando las ecuaciones resumidas en la Tabla 1-2, que están basadas en las definiciones de la Tabla 1-1.

El siguiente paso es la resolución del sistema del carbonato para la temperatura requerida de 25°C, obteniéndose, por tanto, los cuatro parámetros de este sistema (C_T , A_T , pH y pCO_2), con el pH aún en escala *Total* (pH_T). Como último paso, el pH_T obtenido se recalcula para las otras escalas posibles. El resultado de pH en escala SWS se obtiene mediante la aplicación de la primera ecuación de la Tabla 1-2, donde $[SO_4]_T$ es el sulfato total, calculado con las ecuaciones de Morris y Riley (1966), F_T es el fluoruro total, calculado con las ecuaciones de Riley (1965), K_S es la constante de disociación del ion

bisulfato, de Dickson (1990a), K_F es la constante de disociación del ácido fluorhídrico, de Dickson y Riley (Dickson and Riley, 1979) y por último f_H es el coeficiente de actividad del ion hidrógeno en agua de mar (Pérez and Fraga, 1987) calculada de acuerdo con las ecuaciones de Takahashi et al. (1982).

2. Metodología

Los métodos y técnicas de control de calidad que se aplicaron en este capítulo se describen en detalle en Tanhua et al. (2010; 2010b), y se presenta aquí un breve resumen. Esencialmente el procedimiento consiste en la comparación sistemática de datos entre pares de campañas separadas cuyos transectos se cruzan o que tienen estaciones próximas entre ellas. Esto es lo que se denomina análisis mediante *crossovers*. Determinados los desfases (*offsets*) entre pares de campañas, se aplica un procedimiento de inversión para determinar una solución que minimice los desfases proponiendo unos determinados valores de ajuste para las campañas.

Previamente a la comparación es necesario realizar un control de calidad de los datos individuales de cada campaña, el denominado QC1 o 1stQC (Key et al., 2010), que intenta detectar errores producidos en el procesado de los datos y análisis, así como problemas con datos individuales que se marcan como datos anómalos (*outliers*). En este paso se analizan también problemas de excesiva dispersión o inconsistencia con parámetros que presentan patrones de distribución zonal parecida o correlacionables, como los indicados para A_T y pH en la Introducción.

Tras esa fase inicial, se aplicaron los procedimientos de control de calidad secundario (QC2 o 2ndQC) que se detallan en Tanhua et al. (2010; 2010b), y que básicamente consisten en analizar diferencias entre perfiles profundos de estaciones próximas de dos campañas diferentes. Los perfiles se analizan en temperatura potencial, presión y σ_4 (anomalía de la densidad a 4000 m) para intentar descartar la influencia de errores en salinidad y/o temperatura. La profundidad a partir de la cual se analiza cada perfil es variable según las características de la región del *crossover*, para evitar en lo posible la influencia de variaciones estacionales o temporales. En general se emplearon perfiles desde 1500 m hasta fondo en la mayoría de comparaciones. En la Figura 1.5 se reproducen a modo ilustrativo las salidas gráficas de las dos rutinas de software empleadas (Tanhua, 2010; Figuras 2 y 3 de Velo et al., 2010a)

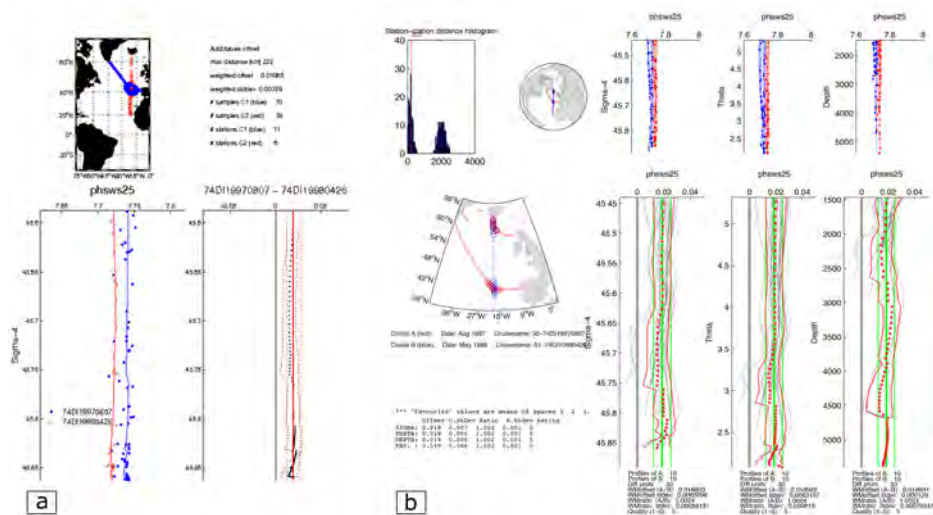


Figura 1.5: Detalle de crossovers generados con las dos rutinas de software empleadas, (a) running-cluster y (b) cnaX, ambas detalladas en Tanhua et al. (2010)

La información relativa al desfase entre cada par de campañas fue analizada y supervisada manualmente por el grupo de CARINA. Para parámetros como la A_T y el pH, el desfase se calculó como desfase aditivo.

Obtenida la información estadística relativa al desfase entre pares de campañas, se realiza el proceso de inversión, mediante la aplicación de un procedimiento de mínimos cuadrados, para obtener una solución de ajustes que minimicen el desfase entre campañas. Los métodos concretos aplicados fueron los denominados “mínimos cuadrados ponderados amortiguado” (WDLSQ, *Weighted Damped Least Square*) y “mínimos cuadrados ponderados” (WLSQ, *Weighted Least Square*). En Tanhua et al. (2010; 2010b) se detallan los procedimientos analizados y la motivación para elegir estos. A diferencia de un ajuste clásico por mínimos cuadrados, los métodos WSLQ y WDSLQ ponderan los desfases entre campañas con la incertidumbre de estos desfases. El método WDSLQ, adicionalmente, amortigua el error del modelo para cada campaña con unas asunciones iniciales de supuesta mejor calidad para ciertas campañas, principalmente secciones largas de campañas WOCE/GLODAP.

3. Umbrales de ajuste

Al igual que para el resto de parámetros, para A_T y pH se definieron unos umbrales mínimos de ajuste, respectivamente, de $6 \mu\text{mol}\cdot\text{kg}^{-1}$ para A_T y de 0.005 unidades para pH, por ser valores ligeramente superiores a la incertidumbre esperable de los métodos de medida. En el proceso de decisión de los ajustes a aplicar por parte del grupo de CARINA, únicamente se tuvieron en cuenta las correcciones obtenidas que fuesen superiores a ese valor.

Las correcciones obtenidas mediante la aplicación de los procedimientos de inversión fueron analizadas individualmente por el grupo de CARINA, que decidió qué correcciones estaban suficientemente justificadas así como el valor de ajuste que se aplicaría sobre el producto final.

4. Evaluación de calidad de datos

Una vez decididos los ajustes para todos los parámetros analizados, aplicados, y generada la base de datos de síntesis que compone el producto principal de CARINA, se realizó un último análisis de crossovers e inversión. Adicionalmente se realizaron determinadas regresiones y análisis estadísticos para asegurar la consistencia de los datos en su región, y la consistencia interna de los parámetros del sistema del carbonato.

Los resultados del último análisis de crossovers efectuado a modo de comprobación final se emplearon también para obtener una estimación de la consistencia interna de los datos de A_T y pH. En la Figura 1.6, reproducida de (Velo et al., 2010a, 2009) se representan los *offsets* obtenidos sobre el producto final de CARINA (i.e. con ajustes aplicados).

Con los datos incluidos en la Figura 1.6 se obtiene, mediante la ecuación (1.1) el nivel promedio de consistencia interna (WM) del producto, que se estima en $3.3 \mu\text{mol}\cdot\text{kg}^{-1}$ para la alcalinidad y 0.0051 unidades para el pH. El valor F indica el porcentaje de *offsets* que son indistinguibles de 0 dada su incertidumbre. L indica el número de crossovers empleados.

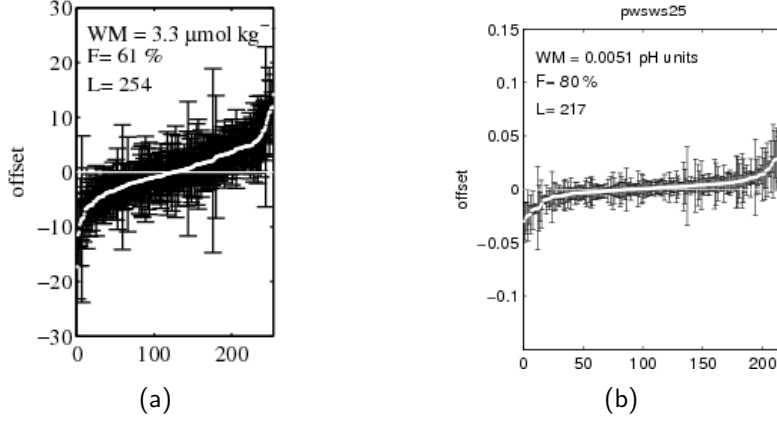


Figura 1.6: *Offsets* obtenidos mediante las rutinas de *crossover* para los conjuntos de datos de CARINA con los ajustes finales ya aplicados, y ordenados de menor a mayor: A_T para CARINA-ATL en el panel (a) y $pH_{SWS 25}$ para el conjunto global de CARINA en el panel (b). WM representa la media ponderada de los *offsets*; F representa el porcentaje de *offsets* indistinguibles de 0 dentro de su incertidumbre y L es el número total de *crossovers* representados.

$$WM = \frac{\sum_{i=1}^L D(i)/\sigma(i)^2}{\sum_{i=1}^L 1/\sigma(i)^2} \quad (1.1)$$

Por último, y para evaluar el nivel general de calidad de datos, se realizaron, tanto para A_T como para pH, análisis mediante regresión multilineal (MLR) en capas de densidad. Se emplearon como parámetros independientes la temperatura potencial (θ), salinidad, latitud, utilización aparente de oxígeno (AOU), nitrato, fosfato y silicato para intentar eliminar la mayor variabilidad natural posible. Estos parámetros se incluyeron conociendo de antemano la interdependencia que existe entre varios de ellos. Sin embargo, el objetivo de este análisis no era generar un modelo estadístico, sino obtener residuos en los que se eliminase, en gran medida, la variabilidad natural de modo que estos residuos pudiesen ser empleados para analizar la calidad de los datos. Este

método de análisis facilita una mejor evaluación de la dispersión y sesgo del conjunto de datos, y es útil para comprobar los ajustes aplicados. Un hecho que debe ser tenido en cuenta es que el procedimiento de MLR puede transmitir el error de las variables explicativas a los residuos de pH, por lo que los resultados deben analizarse conociendo esta limitación.

Con el fin de mejorar la calidad de la evaluación, se aplicó el análisis MLR en cuatro capas de densidad. Se empleó la densidad a 1000 dbar en forma de anomalía (σ_1) para dividir el océano en cuatro capas. La termoclina superior se estableció por $\sigma_1 < 32.25 \text{ kg}\cdot\text{m}^{-3}$. Las aguas intermedias, con profundidades entre 1000 y 2000 metros, se definieron por σ_1 entre el límite anterior y $32.39 \text{ kg}\cdot\text{m}^{-3}$. Las aguas de entre aproximadamente 2000 a 3000 m se definieron mediante los límites de σ_1 de 32.39 y $32.53 \text{ kg}\cdot\text{m}^{-3}$ que corresponden principalmente al Agua Profunda del Atlántico Norte (NADW). Por último la cuarta y última capa aplica a profundidades próximas al fondo, donde domina la presencia de Agua de Fondo Antártica (AABW) y que se definió por valores de $\sigma_1 > 32.53 \text{ kg}\cdot\text{m}^{-3}$. La capa superficial con profundidades de hasta 200 metros no se usó en esta evaluación. Utilizando el análisis MLR, por tanto, se calcularon los residuos de A_T y pH mediante las ecuaciones (1.2):

$$PAR_{MLR} = \sum_{i=1}^8 a_i \cdot X_i \quad (1.2)$$

$$PAR_{residuals} = PAR_{measured} - PAR_{MLR}$$

Donde PAR se refiere al parámetro a calcular (A_T y pH). X_i refiere al valor de cada uno de los predictores empleados (temperatura potencial, salinidad, latitud, AOU, nitrato, fosfato, silicato y un término constante), y a_i a su respectivo coeficiente obtenido mediante el MLR. Este procedimiento se realizó sobre la base de datos de CARINA ya ajustada, y también para la base de datos sin ajustes aplicados.

Los residuos de en cada capa de densidad se muestran en la Figura 1.7 para A_T y Figura 1.8 para pH , reproducidas de sus correspondientes artículos (Velo et al., 2010a, 2009). Las figuras muestran un diagrama de cajas con los residuos para cada parámetro obtenido para cada campaña. Cada caja representa el resumen de cinco números (mínimo, percentil 25, mediana, percentil 75 y máximo). La anchura de las cajas es proporcional al número de muestras de cada campaña, y los números en el eje de abscisas representan el identificador numérico asignado a cada campaña en CARINA (Velo et al., 2010a, 2009).

Para alcalinidad, el mejor ajuste ($R^2=0.95$) se obtiene para las aguas más superficiales, con una desviación estándar promedio de $6.2 \mu\text{mol}\cdot\text{kg}^{-1}$. Las otras capas tienen una desviación estándar ligeramente menor (5.4 y $5.9 \mu\text{mol}\cdot\text{kg}^{-1}$ respectivamente), excepto la capa de fondo, con $6.5 \mu\text{mol}\cdot\text{kg}^{-1}$. La mediana del promedio de desviaciones estándar para todas las campañas es de $4.1 \mu\text{mol}\cdot\text{kg}^{-1}$ para cada una de las capas de densidad más profundas.

Para pH el mejor ajuste de MLR ($R^2 = 0.98$) se obtiene para las aguas más someras, con un error estándar de los residuos de 0.015 . Las siguientes dos capas tienen un error estándar ligeramente inferior (0.012 para las capas 2 y 3 con un R^2 de 0.87 y 0.69 respectivamente), y la capa más profunda tiene

un error estándar de 0.016 (R^2 de 0.77). En términos de desviación media de cada campaña, el error estándar medio de las medianas del conjunto de campañas es de 0.009 unidades de pH para cada una de las tres capas de densidad más profundas.

Los paneles inferiores de cada figura representan los residuos combinados para las cuatro capas. Se puede observar como los residuos son inferiores cuando se utiliza la base de datos corregida en comparación con la original, y que la mayoría de campañas obtienen residuos de A_T y pH dentro de los límites de ajuste.

En este análisis es necesario tener en cuenta también, que adicionalmente a los errores en la medición hay otras fuentes de error en los residuos, como pueden ser los errores en los parámetros que se emplean para la predicción, o que el MLR no sea capaz de explicar toda la variabilidad real del parámetro.

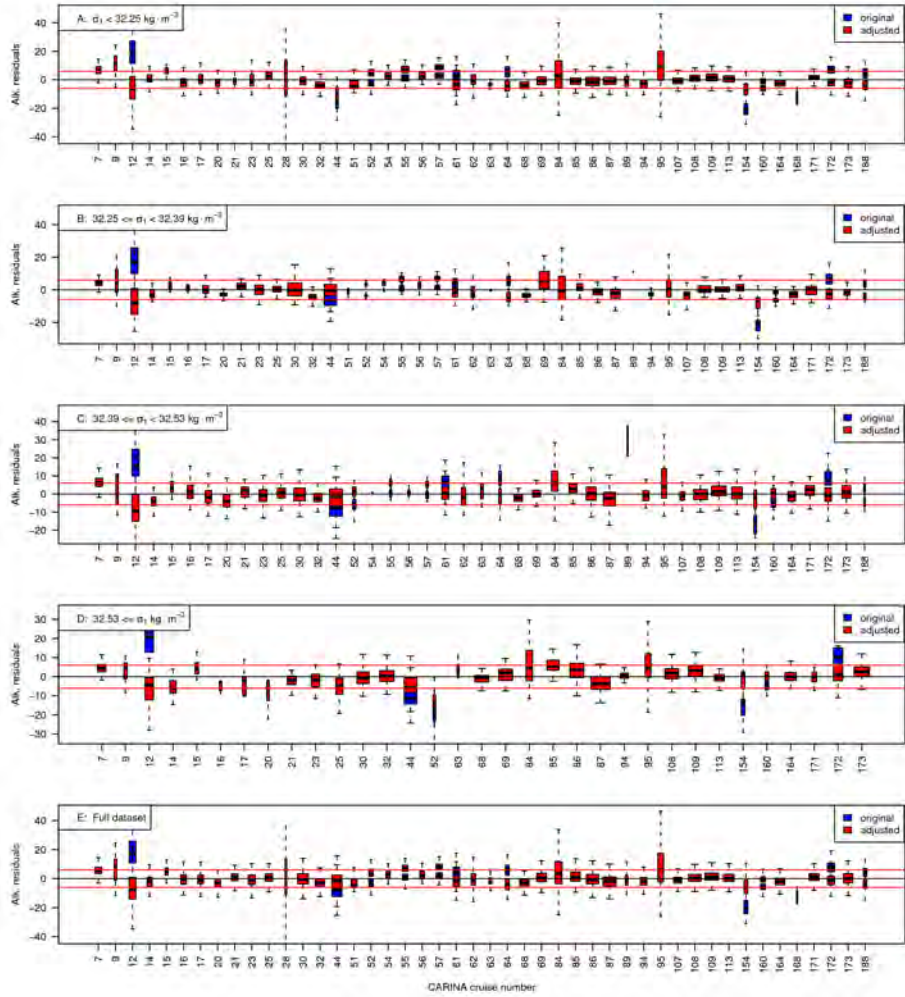


Figura 1.7: Residuos de A_T obtenidos a partir del conjunto de datos de CARINA-ATL aplicando un análisis MLR para alcalinidad frente a temperatura potencial, salinidad, latitud, AOU, nitrato, fosfato y silicato. Los paneles (A) hasta (D) son subconjuntos para los intervalos de σ_1 indicados. El panel (E) representa el conjunto de datos al completo. Los valores azules son los residuos con los datos originales sin ajustar; los valores en rojo son los residuos con los datos ajustados finales. Las líneas rojas representan el límite inferior para ajuste de $\pm 6 \mu\text{mol}\cdot\text{kg}^{-1}$. El ancho de cada caja es proporcional a la cantidad de datos disponibles en cada campaña.

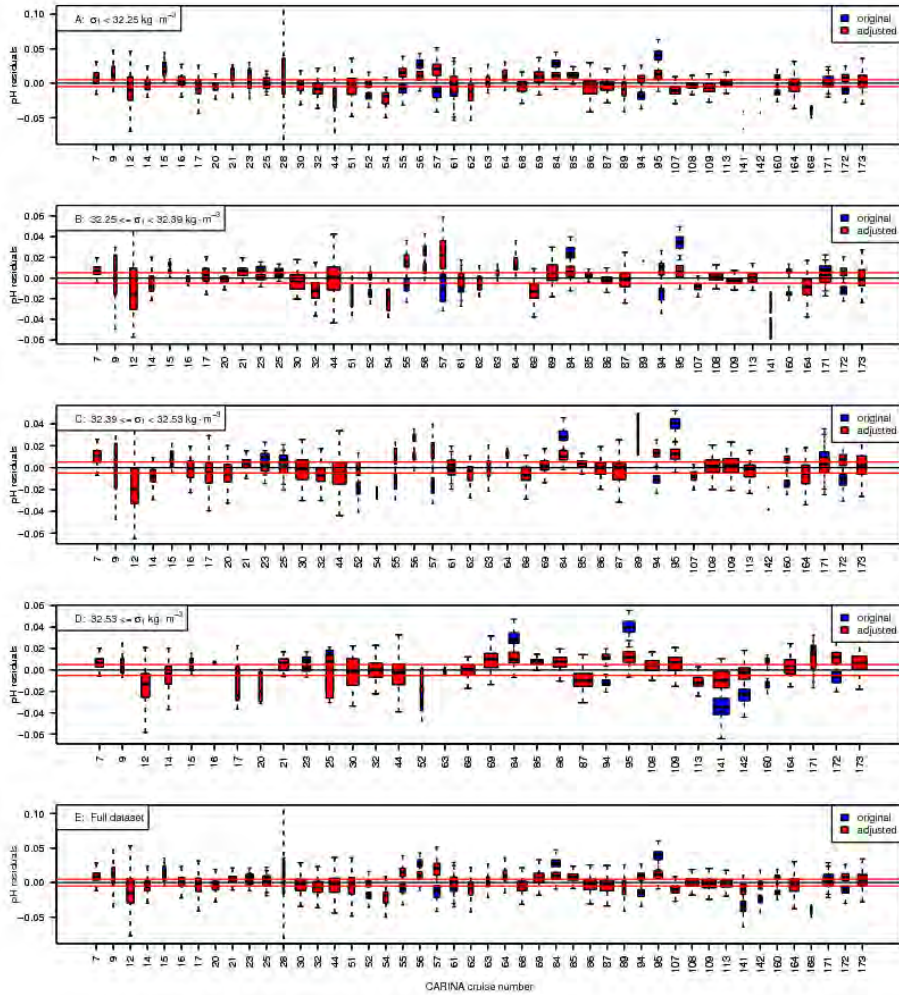


Figura 1.8: Residuos de pH obtenidos a partir del conjunto de datos completo de CARINA aplicando un análisis MLR para alcalinidad frente a temperatura potencial, salinidad, latitud, AOU, nitrato, fosfato y silicato. Los paneles (A) hasta (D) son subconjuntos para los intervalos de σ_1 indicados. El panel (E) representa el conjunto de datos al completo. Los valores azules son los residuos con los datos originales sin ajustar; los valores en rojo son los residuos con los datos ajustados finales. Las líneas rojas representan el límite inferior para ajuste de ± 0.005 unidades de pH. El ancho de cada caja es proporcional a la cantidad de datos disponibles en cada campaña

Capítulo II

Capítulo II – Método de interpolación multiparamétrica

1. Introducción

Este trabajo comenzó también como una contribución al proyecto CARINA, con el objetivo de desarrollar un algoritmo de interpolación que pudiese mejorar la estima geográfica de parámetros en zonas de baja cobertura de datos observacionales y que a su vez fuese fácil de aplicar. Un método de interpolación mejorado podría ser empleado en dicho proyecto para ayudar a obtener un producto general en cuadrícula para los parámetros disponibles. De este modo se facilitaría la generación de una gran base de datos completa del sistema del carbono para el Océano Atlántico partiendo del producto de síntesis de CARINA. Además, el desarrollo podría contribuir a la estimación de inventarios de carbono antropogénico (C_{ant}) de este océano mediante el mallado de los datos de C_{ant} obtenidos por los diversos métodos de estimación disponibles a partir de datos de botella.

Los algoritmos de interpolación desarrollados en el ámbito de este trabajo persiguen el objetivo de ser simples y fáciles de implementar. Después del trabajo pionero de Gandin (1965), introduciendo el análisis objetivo para producir datos en cuadrícula de parámetros meteorológicos de un modo sistemático, los métodos de interpolación objetivos fueron trasladados desde la meteorología

a la oceanografía a finales de los 70s (Bretherton et al., 1976; Freeland and Gould, 1976; Jalickee and Hamilton, 1977). Hoy en día, el análisis objetivo aparece en textos estándar de oceanografía como Bennet (1992) o Emery y Thomson (2001). De hecho, una de las bases de datos utilizadas en el presente estudio, el Atlas Mundial del Océano de 2005 (WOA05), se elaboró empleando estas técnicas de análisis. El presente estudio utiliza un algoritmo multiparamétrico de distancias inversas que se aplicó a los datos del producto final de CARINA (Key et al., 2010; Tanhua et al., 2009), y empleó los datos interpolados objetivamente de WOA05 como malla de referencia para calcular las distancias multiparamétricas.

El proyecto CARINA ha completado su base de datos empleando únicamente campañas en las cuales se midieron parámetros del sistema del carbono, por lo que la cobertura de observaciones es baja en algunas regiones como el Océano Austral. En este contexto, un método de interpolación basado únicamente en distancias geográficas podría no funcionar bien debido precisamente a la poca densidad de datos. Un posible modo de aliviar ese problema consiste en intentar incorporar más información en el algoritmo de interpolación que no sea únicamente espacial. La incorporación de propiedades conservativas de las diferentes masas de agua puede ofrecer un beneficio, ya que proporciona un mejor ajuste a las distribuciones reales que un método basado en la distancia puramente espacial. Una ventaja adicional de este procedimiento es que los artefactos que pueden aparecer en las distribuciones de las masas de agua derivados de interpolaciones espaciales sencillas pueden evitarse más fácilmente.

El producto de datos de CARINA no se distribuye sobre una malla uniforme sino que está compuesto principalmente de estaciones de CTD dispersas, organizadas en secciones transoceánicas. En términos de parámetros registrados, el conjunto de datos compila muchos parámetros bioquímicos incluyendo los necesarios para la estimación de carbono antropogénico por diferentes métodos. Por el contrario, el conjunto de datos de WOA05 está estructurado en una malla tridimensional homogénea con parámetros termohalinos y biogeoquímicos definidos en los nodos de la malla, pero carece de muchos de los parámetros necesarios para los cálculos del sistema del carbonato. Por tanto, la generación de un algoritmo que combina las propiedades del conjunto de datos de WOA05 y CARINA parece la forma adecuada de proceder. Por último, la aplicación del método de interpolación sobre una estimación de carbono antropogénico proporcionará datos de C_{ant} interpolados sobre la malla estándar de WOA, permitiendo así el tratamiento común de la información hidrográfica disponible en ambos conjuntos de datos.

Como modo de evaluar los resultados, se compararon dos variantes del método interpolación, una basado únicamente en distancias espaciales y otra que utiliza trazadores físicos y biogeoquímicos (referida en adelante como método de interpolación de masas de agua (WMP)). El contraste del comportamiento individual de los dos métodos se realizó empleando parámetros dependientes (interpolando parámetros incluidos en las distancias multiparamétricas) y un parámetro independiente (oxígeno, no incluido en las distancias multiparamétricas).

Como las comprobaciones de calidad del método de interpolación WMP ofrecieron resultados positivos, se dio un paso hacia delante: interpolar el carbono antropogénico sobre la cuadrícula de WOA. El importante papel que desempeñan los océanos en el ciclo global del carbono es indiscutible, ya que tienen la capacidad de secuestrar 2.2 ± 0.4 Pg-C al año, lo que representa un 25 % del total de carbono antropogénico emitido a la atmósfera (8.0 ± 0.5 Pg-C año⁻¹) (Canadell et al., 2007). Más notablemente, el Océano Atlántico almacena el 38 % del carbono antropogénico oceánico (Sabine et al., 2004) aunque representa el 29 % del área global de la superficie oceánica. La particular dinámica del Océano Atlántico permite la formación de aguas profundas en el Atlántico Norte y esto incrementa los flujos de captación y almacenamiento de carbono antropogénico de esta cuenca. Procesos recientemente detectados provocados por cambios decenales del clima mundial, como la desaceleración de la circulación termohalina, parecen haber contribuido de manera significativa a la reducción de la capacidad de almacenamiento de C_{ant} en el Atlántico Norte y el Océano Austral (Joos et al., 1999; Le Quéré et al., 2007). La yuxtaposición de estos efectos opuestos ha incrementado dramáticamente la necesidad de estimar con precisión el estado de los inventarios de C_{ant} y ha elevado la importancia de afinar los métodos de interpolación aplicados a conjuntos de datos dispersos.

Se emplearon dos métodos para obtener la estimación de C_{ant} sobre la totalidad del Océano Atlántico, el φC_T^0 y el TrOCA. El método de estimación φC_T^0 fue elegido por ser desarrollado por coautores (Vázquez-Rodríguez et al., 2009a), haber participado en su implementación (elaboración de paquete

Matlab, disponible en: <http://oceano.iim.csic.es/co2group/>) y por ser sencillo de aplicar y verificar. También es un método actualizado y funciona comparativamente bien con respecto a otros métodos (Vázquez-Rodríguez et al., 2009b). El método TrOCA (Touratier et al., 2007) fue considerado adicionalmente como una referencia de apoyo debido a la facilidad de su aplicación. El carbono antropogénico fue computado aplicando estos métodos de estimación al conjunto de datos de CARINA y, posteriormente, interpolado empleando dos versiones de los algoritmos de interpolación, uno puramente espacial y el multiparamétrico WMP. El siguiente paso fue calcular los volúmenes para obtener los inventarios. Para dicho cálculo se empleó ETO-PO2v2 (U.S. Department of Commerce, 2006) como referencia para el fondo oceánico.

2. Metodología

El método de interpolación desarrollado en el presente estudio presenta dos características que lo diferencian de otros métodos clásicos. Uno es el hecho de emplear ventanas móviles espaciales, y el otro es el empleo de distancias relativas en el espacio definido por varios parámetros de las masas de agua en vez o adicionalmente a las distancias geográficas clásicas.

2.1. Ventana móvil espacial

El concepto de ventana móvil espacial se refiere a la selección de ventanas o cajas de datos que rodean el nodo de la malla que queremos interpolar. Es decir, la ventana define los límites espaciales con los que se seleccionarán

los datos originales que se usarán para calcular el valor del nodo a interpolar. Estas ventanas pueden ser de tamaño fijo o variable pero son únicas para cada nodo que se quiere interpolar, por lo que se definen y construyen alrededor de cada nodo, y el algoritmo de interpolación recorre iterativamente todos los nodos a interpolar.

Para la geometría de la ventana se prefirió emplear un cubo con centro en el nodo frente a una forma esférica que sería idealmente más adecuada. El motivo de esta decisión fue puramente de rendimiento computacional.

En cuanto a las dimensiones horizontales de la ventana, se optó por establecer un tamaño inicial de $2^\circ \times 2^\circ$ (latitud x longitud), incrementando el tamaño a $10^\circ \times 10^\circ$ en caso de que se encontrasen menos de 20 muestras en la ventana más pequeña, y hasta $20^\circ \times 20^\circ$ en caso de que tampoco se encontrasen 20 muestras en la ventana intermedia. Para la dimensión vertical se empleó un intervalo en metros de $\pm(150 + 0.1 \times \text{profundidad del nodo})$. Esta definición de ventana permite evitar suavizados espaciales excesivos o pérdida de señal al restringir, cuando sea posible, la interpolación a un área próxima, pero al mismo tiempo permite obtener resultados en áreas en las que los datos están dispersos.

2.2. Distancias multiparamétricas

Las distancias multiparamétricas se refieren al empleo de normas euclídeas tomadas en espacios multiparamétricos y no en espacios solamente geográficos. Son distancias tomadas en coordenadas, además de geográficas, definidas por parámetros conservativos que permiten identificar o trazar masas de agua.

De este modo, para asignar un valor interpolado a un nodo, el método otorga mayor peso a las muestras (dentro de la ventana) más cercanas en salinidad, temperatura potencial, 'NO' o 'PO' (Broecker and Peng, 1982; Broecker, 1974; Pérez et al., 1993; Ríos et al., 1989). Esto, junto con la incorporación de una climatología como la de WOA05, que ofrece estos parámetros conservativos en una malla distribuida regularmente y a una resolución que mejora la de las muestras discretas de CARINA, permite incorporar la información de la distribución de las estructuras de masas de agua a las muestras discretas que queremos interpolar, ofreciendo un resultado más acorde a la realidad que el que podría ofrecer un método puramente geográfico.

2.3. Algoritmo base

Como raíz del método se emplea un esquema clásico de interpolación por distancias inversas, en el que los valores más alejados (en distancia solamente geográfica o multiparamétrica) influyen en menor medida en el valor interpolado final. Se emplea la ecuación (2.1) como base para el método. En ella se representa como PAR al parámetro que se pretende interpolar. Las muestras de la base de datos sin estructura geográfica homogénea (CARINA) se representan por el subíndice i , y los nodos de la malla (WOA) como j . El factor f se define en la ecuación (2.2) y representa el factor de distancias que se aplica a la inversa $\left(f_i^j\right)^{-1}$, obteniendo así el valor interpolado para el nodo por promedio ponderado.

$$PAR_j = \frac{\sum_i PAR_i \cdot (f_i^j)^{-1}}{\sum_i (f_i^j)^{-1}} \quad (2.1)$$

Los factores de ponderación (f_i^j) por distancia para cada dato origen se calculan mediante el uso del cuadrado de las distancias multiparamétricas. De este modo, para cada muestra (i) de CARINA que se emplee para obtener un nodo (j) en la cuadrícula de WOA05, se calculará un factor de distancia al nodo que se define del siguiente modo:

$$\begin{aligned} f_i^j = & w_{lat} \left(\frac{lat^i - lat^j}{\Delta lat} \right)^2 + w_{lon} \left(\frac{lon^i - lon^j}{\Delta lon} \right)^2 + w_z \left(\frac{z^i - z^j}{\Delta z} \right)^2 \\ & + w_{\theta} \left(\frac{\theta^i - \theta^j}{std(\theta^j)} \right)^2 + w_S \left(\frac{S^i - S^j}{std(S^j)} \right)^2 \\ & + w_{NO} \left(\frac{NO^i - NO^j}{std(NO^j)} \right)^2 + w_{PO} \left(\frac{PO^i - PO^j}{std(PO^j)} \right)^2 \end{aligned} \quad (2.2)$$

La ecuación (2.2) muestra el algoritmo genérico de cálculo de factores por distancias, que emplea tanto parámetros de distancias geográficas (latitud: lat , longitud: lon , profundidad: z) como distancias paramétricas en los espacios de temperatura potencial, salinidad, 'NO' y 'PO' (θ , S , NO , PO). La aportación al factor de ponderación por parte de cada parámetro se controla con los pesos w_x (x : lat , lon , z , θ , S , NO , PO), a los que en el presente estudio se ha asignado un valor de 1 o 0 según el interés en incorporar cada parámetro, pero que admiten cualquier otra combinación de valores numéricos que el investigador quiera emplear. Además, para que las distancias parciales

$(x^i - x^j)$ que cada parámetro aporta al factor por distancia sean comparables y sumables, dichas distancias parciales se hacen adimensionales dividiendo por las desviaciones estándar ($std(x^j)$) del conjunto de los nodos WOA, que caen dentro de la ventana, o por las dimensiones de la ventana (Δx), según corresponda.

De esta manera, el factor por distancia calculado con la ecuación anterior para cada dato de CARINA se puede interpretar como la suma de las distancias cuadráticas de sus parámetros con respecto a las distribuciones normales con el valor del nodo como media y desviaciones estándar las estimadas por los datos WOA en la ventana. Al utilizar una norma L^2 (cuadrática) para calcular los factores finales, se acentúa de forma no lineal la importancia de las distancias parciales mayores, i.e. la influencia de los parámetros más alejados de su distribución normal de referencia es mayor (en el factor, resultando inversa en el valor interpolado).

Para diferenciar los métodos Espacial (2.3) y WMP (2.4) , tal como se indicaba en el párrafo anterior, se asignan pesos diferentes a los factores individuales, eligiéndose las siguientes combinaciones para este estudio:

$$w_{lat}, w_{lon}, w_z, w_\theta, w_S, w_{NO}, w_{PO} = (1, 1, 1, 0, 0, 0, 0) \quad (2.3)$$

Representa un esquema puramente espacial en el que a todas las coordenadas espaciales se les asigna el mismo peso, y las paramétricas se eliminan de la ecuación.

$$w_{lat}, w_{lon}, w_z, w_\theta, w_S, w_{NO}, w_{PO} = (0, 0, 0, 1, 1, 1, 1) \quad (2.4)$$

Representa un esquema puramente multiparamétrico, ya que se elimina el uso de distancias espaciales y sólo se tiene en cuenta la similitud a nivel de parámetros conservativos de las masas de agua.

El software a través del cual se implementa el algoritmo permite la interpolación conjunta de varios parámetros en la misma ejecución, ya que para conjuntos de datos completos, los factores de distancias son los mismos para los diferentes parámetros a interpolar. De este modo se evita el recómputo de los factores, lo que permite interpolar simultáneamente un gran número de parámetros con un coste extra muy pequeño en cuanto a tiempos de computación. Esto es útil para trabajos como el realizado en CARINA ya que permite la generación del producto en cuadrícula WOA de todos los parámetros en un tiempo menor. A modo de valor orientativo del orden de magnitud de los tiempos de cálculo, la interpolación de ≈ 50000 valores de C_T de la base de datos de CARINA a ≈ 280000 nodos de la malla de WOA para el Océano Atlántico se realiza aproximadamente en unos 20 minutos en un equipo Dual Intel Xeon E5620 del año 2010

3. Evaluación

Para evaluar la calidad de ambos tipos de interpolaciones (Espacial y WMP), se evaluaron en primer lugar los resultados de la interpolación de temperatura potencial, salinidad, 'NO', y 'PO' de la base de datos de botella del producto de CARINA sobre la cuadrícula de WOA05 contra, precisamente, los datos de la propia climatología de WOA05 para esos parámetros. A continuación, y dado que estos parámetros están involucrados en el cálculo de los factores de

la interpolación, se realizó una última evaluación más independiente comparando el oxígeno interpolado a partir de los datos de CARINA contra el de la climatología de WOA05.

Para cada parámetro y método de interpolación se sustrajeron, nodo a nodo, los valores de referencia equivalentes de WOA05, para obtener matrices de anomalías de la interpolación con respecto a WOA05. En esas matrices de anomalías se calculó la media y la desviación estándar del conjunto de los datos, y de determinados subdominios espaciales y en función de la temperatura potencial. Se caracterizó también, con el coeficiente R^2 , la correlación entre los datos interpolados y los de referencia para cada dominio. Los dominios seleccionados geográficamente fueron tres en función de la variabilidad general de las masas de agua, con las latitudes al norte de 30°N , la zona tropical, y las latitudes al sur de los 30°S . En profundidad, se establecieron dos niveles en función de la temperatura potencial, estableciendo el valor de separación en 5°C , ya que la isoterma de 5° representa una frontera aproximada entre las aguas profundas poco ventiladas y las superiores e intermedias más ventiladas. Esta división en profundidad también divide los inventarios de C_{ant} aproximadamente a la mitad. En la Tabla II-1 se muestran los resultados obtenidos.

En la Figura 2.1 se muestra la media anual de la temperatura potencial de la climatología de WOA05 a lo largo de una sección sobre el meridiano 28°W . La temperatura potencial interpolada espacialmente se asemeja bastante a primera vista, pero se observan varios desajustes cuando se analizan detalladamente los residuos. Los mayores desajustes aparecen en la capa su-

Total	R ²	Total Mean	Std	R ²	$\theta > 5^{\circ}\text{C}$ Mean	Std	R ²	$\theta < 5^{\circ}\text{C}$ Mean	Std
θ_{spatial}	0.964	0.34	1.02	0.856	0.23	0.62	0.914	0.55	1.48
θ_{WMP}	0.993	0.10	0.43	0.952	0.11	0.37	0.988	0.07	0.53
S_{spatial}	0.925	0.01	0.14	0.887	-0.02	0.07	0.905	0.04	0.22
S_{WMP}	0.984	0.00	0.07	0.954	-0.01	0.05	0.981	0.00	0.09
$\text{NO}_{\text{spatial}}$	0.948	0.78	18.80	0.845	3.78	15.46	0.918	-4.70	22.71
NO_{WMP}	0.989	2.55	8.65	0.957	3.57	8.16	0.983	0.44	9.22
$\text{PO}_{\text{spatial}}$	0.956	-5.66	18.56	0.893	-4.29	15.29	0.919	-8.17	23.18
PO_{WMP}	0.991	-2.98	8.35	0.971	-3.13	8.15	0.984	-2.67	8.75
$\text{O}_{2\text{spatial}}$	0.901	4.14	14.47	0.857	4.40	13.12	0.918	3.65	16.64
$\text{O}_{2\text{WMP}}$	0.943	2.65	10.96	0.945	2.67	8.36	0.921	2.60	14.95
Lat > 30° N									
θ_{spatial}	0.934	0.20	1.24	0.860	0.34	1.61	0.802	0.01	0.42
θ_{WMP}	0.987	0.06	0.55	0.987	0.06	0.55	0.818	0.04	0.41
S_{spatial}	0.832	0.02	0.20	0.752	0.03	0.27	0.716	0.00	0.06
S_{WMP}	0.948	0.01	0.11	0.948	0.01	0.11	0.874	0.01	0.09
$\text{NO}_{\text{spatial}}$	0.954	-0.96	12.72	0.916	-3.58	15.69	0.637	2.26	6.27
NO_{WMP}	0.988	0.71	6.56	0.988	0.71	6.56	0.819	2.44	4.34
$\text{PO}_{\text{spatial}}$	0.937	-5.06	15.58	0.877	-6.26	19.55	0.574	-3.60	8.20
PO_{WMP}	0.988	-2.15	6.97	0.988	-2.15	6.97	0.821	-1.64	4.82
$\text{O}_{2\text{spatial}}$	0.928	0.59	9.26	0.875	0.14	11.19	0.877	1.14	6.06
$\text{O}_{2\text{WMP}}$	0.946	-0.61	8.20	0.946	-0.61	8.20	0.909	0.49	5.43
Tropical									
θ_{spatial}	0.982	0.30	0.83	0.958	0.59	1.11	0.985	0.03	0.14
θ_{WMP}	0.998	0.06	0.27	0.995	0.11	0.38	0.994	0.02	0.08
S_{spatial}	0.956	0.01	0.12	0.943	0.03	0.17	0.981	-0.01	0.02
S_{WMP}	0.995	0.00	0.04	0.994	0.00	0.06	0.993	0.00	0.01
$\text{NO}_{\text{spatial}}$	0.981	1.59	10.99	0.968	-1.69	13.62	0.898	4.70	6.31
NO_{WMP}	0.995	2.52	5.34	0.992	1.78	6.61	0.964	3.18	3.75
$\text{PO}_{\text{spatial}}$	0.979	-2.19	11.57	0.967	-5.12	13.44	0.858	0.59	8.60
PO_{WMP}	0.995	-1.02	5.72	0.992	-1.65	6.50	0.955	-0.46	4.85
$\text{O}_{2\text{spatial}}$	0.926	5.69	14.06	0.831	7.96	19.01	0.970	3.55	5.76
$\text{O}_{2\text{WMP}}$	0.937	5.66	12.79	0.860	6.60	17.24	0.963	4.82	6.49
Lat < 30° S									
θ_{spatial}	0.916	0.44	1.05	0.821	0.77	1.93	0.741	0.38	0.75
θ_{WMP}	0.983	0.14	0.47	0.969	-0.04	0.66	0.920	0.17	0.43
S_{spatial}	0.768	0.00	0.13	0.820	0.09	0.22	0.747	-0.02	0.09
S_{WMP}	0.958	-0.01	0.05	0.968	-0.01	0.09	0.941	-0.01	0.05
$\text{NO}_{\text{spatial}}$	0.817	0.88	24.89	0.822	-13.49	39.92	0.527	3.68	19.54
NO_{WMP}	0.964	3.36	11.10	0.957	-1.23	15.52	0.855	4.03	10.13
$\text{PO}_{\text{spatial}}$	0.846	-8.69	23.18	0.829	-18.21	38.51	0.646	-6.83	18.23
PO_{WMP}	0.971	-4.90	10.07	0.968	-5.82	13.20	0.895	-4.76	9.52
$\text{O}_{2\text{spatial}}$	0.750	4.44	16.26	0.775	-1.44	14.47	0.732	5.58	16.35
$\text{O}_{2\text{WMP}}$	0.908	1.65	9.79	0.834	-1.71	11.37	0.916	2.14	9.44

Tabla II-1: Coeficiente de determinación (R^2), media (Mean) y desviación estándar (Std) resultantes de sustraer los valores de WOA05 a los datos interpolados. Los datos están computados para la cuenca completa del Océano Atlántico, para cada subdominio y para ambas posibilidades de interpolación (Espacial y WMP). La temperatura se indica en $^{\circ}\text{C}$, Salinidad en psu, y 'NO', 'PO', O_2 en $\mu\text{mol}\cdot\text{kg}^{-1}$.

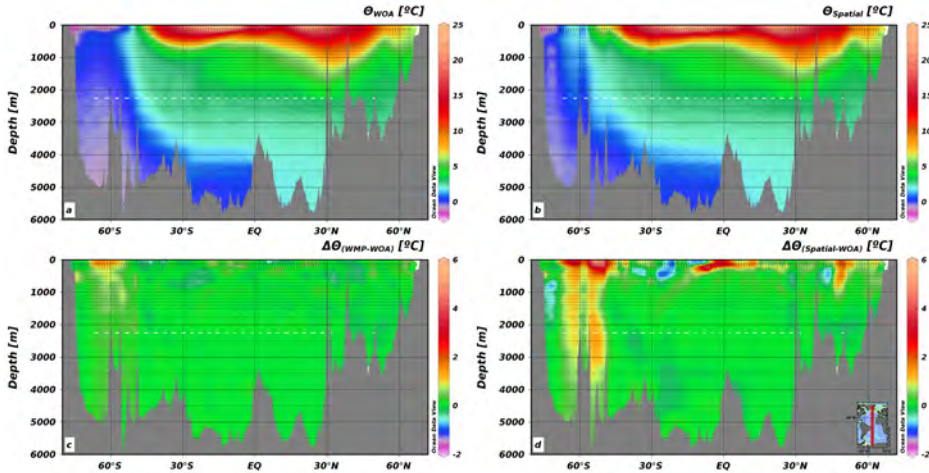


Figura 2.1: Variabilidad de la temperatura potencial ($^{\circ}\text{C}$) a lo largo de la sección 28W. El panel (a) muestra los datos de WOA05; el panel (b) los datos interpolados espacialmente a partir de CARINA; el panel (c) muestra los residuos de la interpolación WMP respecto a WOA05 y el panel (d) los residuos de la interpolación espacial respecto a WOA05.

terior, con diferencias de hasta 2°C en valor absoluto. En las aguas profundas hay una mejor concordancia, aunque aparece una ancha banda al sur de los 40°S con un sesgo sistemático superior a 1°C . Por otro lado, analizando los residuos de la interpolación WMP (θ_{WMP} panel c) se observa que ofrecen un mejor ajuste y menores diferencias en valor absoluto que el método puramente espacial. El mejor ajuste se obtiene también para aguas profundas al norte de los 45°S . Estadísticamente (Tabla II-1), el mejor ajuste se obtiene mediante la interpolación WMP independientemente de si se analiza la interpolación para todo el Atlántico o para los subdominios.

Las mayores diferencias entre la interpolación espacial y la WMP se observan en la capa superior ($\theta > 5^{\circ}\text{C}$) donde los tres estadísticos (Tabla II-1) del ajuste mejoran notablemente con la interpolación WMP. Más específicamente, los mejores ajustes (R^2) se obtienen en la región tropical y latitudes

norte, mientras que los peores ajustes se obtienen en latitudes sur. Esto se debe, probablemente, a la menor densidad de datos de CARINA en el Océano Austral.

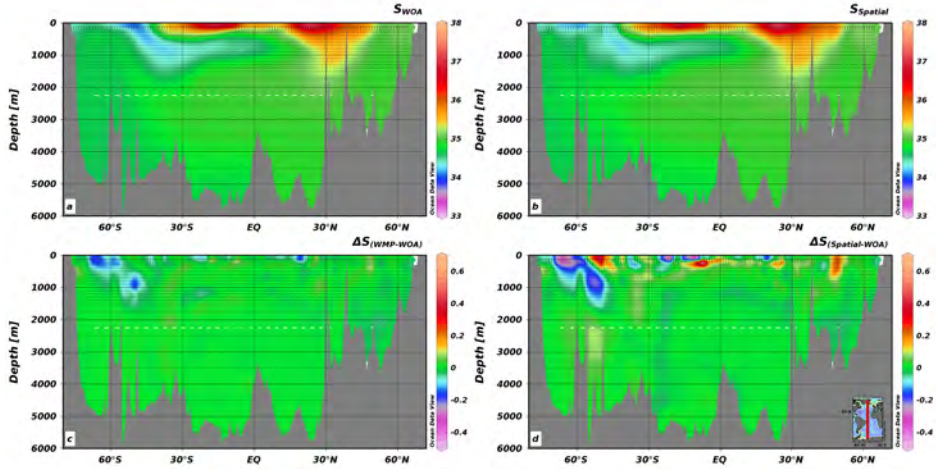


Figura 2.2: Variabilidad de la salinidad (psu) a lo largo de la sección 28W. El panel (a) muestra los datos de WOA05; el panel (b) los datos interpolados espacialmente a partir de CARINA; el panel (c) muestra los residuos de la interpolación WMP respecto a WOA05 y el panel (d) los residuos de la interpolación espacial respecto a WOA05.

La Figura 2.2 muestra la media climatológica anual de salinidad a lo largo de la sección 28°W. Como en el caso de la temperatura potencial (Figura 2.1), el ajuste de la interpolación espacial es bueno en general, pero se pueden apreciar errores considerables alrededor del mínimo de salinidad, asociado con la presencia de las aguas intermedias antárticas (AAIW) (Mémery et al., 2000). Los mayores errores aparecen también nuevamente en aguas superficiales, con desajustes superiores a 0.2 psu. Aunque en aguas profundas la concordancia es alta en general, al sur de los 40°S siguen apareciendo zonas con grandes sesgos al igual que para la interpolación espacial de θ . La interpolación WMP de la salinidad muestra también una mejor concordancia, con residuos más

bajos. Al igual que para θ , la mejor concordancia se obtiene para aguas al norte de los 40°S.

Las mayores discrepancias obtenidas entre los resultados de ambos métodos se encuentran en las capas superiores ($\theta > 5^\circ\text{C}$) del Océano Atlántico. En este dominio, los R^2 obtenidos por el método WMP son mucho mejores que los producidos por la interpolación espacial. Sin embargo, los índices R^2 obtenidos para la salinidad son ligeramente menores que los de temperatura potencial.

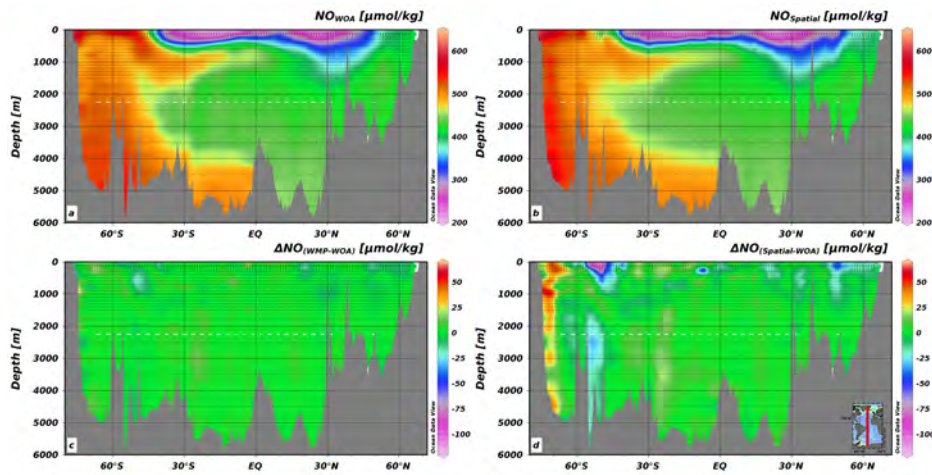


Figura 2.3: Variabilidad del 'NO' ($\mu\text{mol}\cdot\text{kg}^{-1}$) a lo largo de la sección 28W. El panel (a) muestra los datos de WOA05; el panel (b) los datos interpolados espacialmente a partir de CARINA; el panel (c) muestra los residuos de la interpolación WMP respecto a WOA05 y el panel (d) los residuos de la interpolación espacial respecto a WOA05.

En cuanto a las interpolaciones de 'NO' y 'PO', están muy correlacionadas, por lo que se ha incluido únicamente la Figura 2.3 con el parámetro 'NO'. Se puede destacar que, al igual que para salinidad o temperatura potencial, las interpolaciones espaciales se asemejan bastante a los datos de WOA05, con R^2 próximos a 0.95 para el Océano Atlántico Norte en conjunto. Las capas superiores, en particular las del Océano Austral muestran las mayores

anomalías, con desfases de hasta $50 \mu\text{mol}\cdot\text{kg}^{-1}$ para 'NO'. La interpolación WMP, muestra un mejor ajuste en general, con menor desviación estándar y mayores coeficientes de determinación en cada subdominio. Al contrario que el caso de temperatura potencial, el mejor ajuste se obtiene esta vez en las capas más cálidas, lo que se interpreta como que es debido a la relativa alta variabilidad de 'NO' y 'PO' en las capas superiores. Por el contrario, el peor ajuste se obtiene en las capas profundas del Océano Austral.

3.1. Oxígeno

Como se comentó anteriormente, la interpolación de Oxígeno supone una buena prueba de la calidad de las interpolaciones, ya que la interpolación multiparamétrica emplea únicamente distancias en parámetros conservativos, y el oxígeno está disponible en ambos conjuntos de datos.

A modo de introducción, podemos destacar que las concentraciones de oxígeno están controladas por las bombas biológicas y de solubilidad. Si nos fijamos en la sección 28°W de la Figura 2.4, podemos observar el mínimo de las capas superiores de la región tropical provocado por la remineralización de la materia orgánica (predominio de la bomba biológica). Por el contrario, los valores altos observados en las zonas polares son consecuencia de la alta solubilidad de oxígeno en las aguas frías superficiales de estas regiones (la bomba de solubilidad prevalece en este dominio) y las altas tasas de renovación.

La interpolación espacial de los datos de oxígeno de CARINA, muestra una distribución parecida a la de WOA05, con un R^2 de 0.90 para el conjunto del Océano Atlántico. En detalle, sin embargo, se puede observar la tendencia

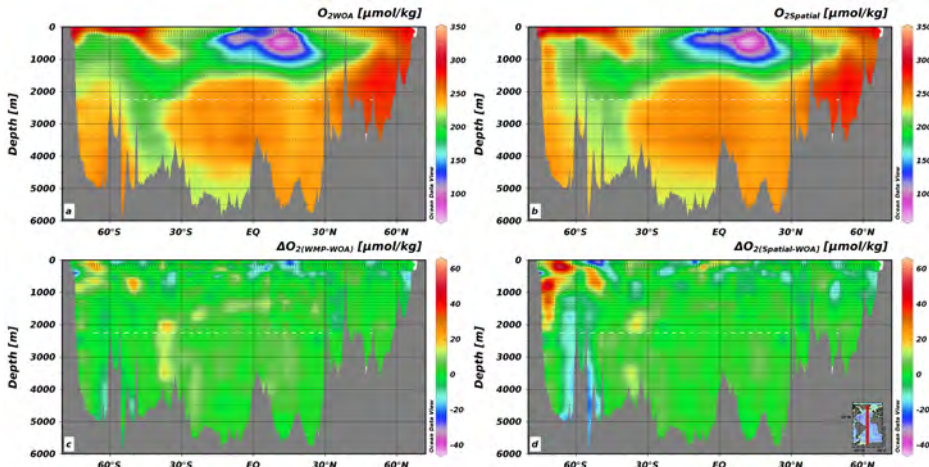


Figura 2.4: Variabilidad de O_2 disuelto ($\mu\text{mol}\cdot\text{kg}^{-1}$) a lo largo de la sección 28W. El panel (a) muestra los datos de WOA05; el panel (b) los datos interpolados espacialmente a partir de CARINA; el panel (c) muestra los residuos de la interpolación WMP respecto a WOA05 y el panel (d) los residuos de la interpolación espacial respecto a WOA05.

a acentuar algunos gradientes en latitudes sur. Las mayores diferencias se observan en las capa superior y en el Océano Austral, con desajustes que pueden llegar hasta $50 \mu\text{mol}\cdot\text{kg}^{-1}$. Los mejores ajustes, se obtienen en las aguas frías profundas de la región tropical.

La interpolación WMP del oxígeno de CARINA muestra un mejor ajuste con la climatología de WOA, con menores residuos y mejor R^2 en general que la interpolación espacial. Al contrario que en el resto de dominios, en las aguas profundas tropicales, la interpolación WMP parece no mostrar todo su potencial, y los resultados de la interpolación espacial son más acordes con la climatología de WOA. Sin embargo, en el Océano Austral, donde la cobertura de datos es escasa, la interpolación WMP es donde ofrece unas estimaciones más robustas.

3.2. Carbono antropogénico.

Una vez evaluados ambos métodos de interpolación, se realizó, a modo de caso de aplicación, una interpolación de las estimaciones de C_{ant} realizadas con los métodos φC_T^0 y TrOCA (Touratier et al., 2007; Vázquez-Rodríguez et al., 2009b; Velo et al., 2010b) partiendo del producto de CARINA, para generar un producto en cuadrícula estándar de carbono antropogénico en el Océano Atlántico, que permite analizar la distribución espacial y calcular inventarios.

Dado que el C_{ant} es indistinguible del CO_2 natural, no hay referencias absolutas contra las que comparar. La Figura 2.5 muestra las interpolaciones Espacial y WMP para los métodos φC_T^0 y TrOCA. Se puede observar que el patrón general de distribución es similar al dado por Lee et al. (2003) y Vázquez-Rodríguez et al. (2009b). Sabine et al. (2004) calculó un inventario total para el Atlántico de 40Pg-C empleando la base de datos de GLODAP interpolada en la cuadrícula de WOA. Previamente Lee et al. (2003) había obtenido un inventario de 47Pg-C empleando la misma base de datos pero sin interpolar a cuadrícula. Empleando datos de CFC, Waugh et al. (2004) obtuvo un inventario total de 48 Pg-C. Un estudio comparativo realizado por Vázquez-Rodríguez et al. (2009b) empleando cinco campañas transoceánicas largas obtuvo inventarios de 55 y 51 Pg-C para los métodos φC_T^0 y TrOCA respectivamente.

Los valores que se obtuvieron en este trabajo empleando la base de datos de CARINA interpolada a la red de WOA fueron similares para el método φC_T^0 (55 Pg-C), pero algo mayores para TrOCA (58 Pg-C).

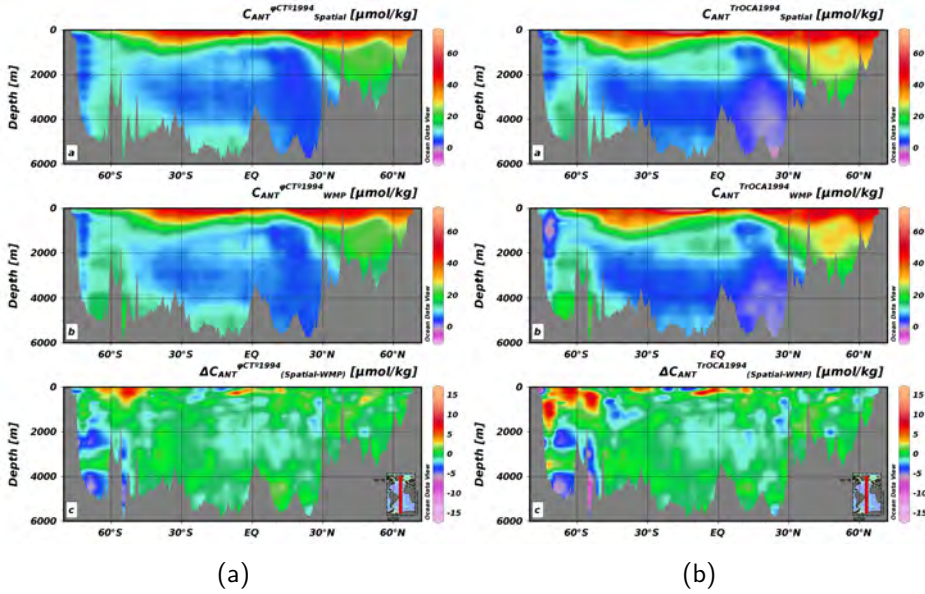


Figura 2.5: Estimaciones de $C_{ant}(\mu\text{mol}\cdot\text{kg}^{-1})$ por los métodos φC_T^0 (Columna izquierda) y TrOCA (Columna derecha) a lo largo de la sección 28W empleando la interpolación espacial (paneles a) y la WMP (paneles b). En los paneles (c) se muestran las anomalías entre ambos métodos (Espacial – WMP)

Las principales diferencias con respecto a los datos en cuadrícula obtenidos a partir de GLODAP (Key et al., 2004; Lee et al., 2003) se observan en el Océano Atlántico Sur, donde un gran número de estimaciones de GLODAP, obtenidas a partir del método ΔC^* (Gruber et al., 1996), son negativas. Los valores de C_{ant} de GLODAP para el Océano Austral son inferiores de los computados aquí.

Capítulo III

Capítulo III – Estimación de A_T por métodos MLR y neuronales

Durante la última década se han puesto a disposición del público dos colecciones importantes de datos hidrográficos relevantes al sistema del carbono: GLODAP y CARINA. Estas dos colecciones están compuestas por una síntesis efectuada sobre datos de botella para todas las profundidades del océano y para un gran número de campañas oceanográficas (GLODAP global: 115 campañas, CARINA global: 188 campañas) realizadas a lo largo de varias décadas. En la mayoría de las campañas se midieron al menos dos parámetros del sistema del carbonato, pero existen un gran número de estaciones, muestras o incluso campañas para las cuales el sistema del carbonato está infradeterminado, (es decir sólo uno o ningún parámetro de este sistema fue medido) generándose, por tanto, una falta de datos para el sistema del carbonato en muestras de estas colecciones. Un método para llenar estos huecos sería muy útil, ya que podría ayudar en las estimaciones del contenido de carbono antropogénico (C_{ant}) o la acidificación oceánica. El objetivo de este trabajo es elaborar y aplicar un algoritmo de regresión multilineal (MLR) en ventana móvil 3D para completar los huecos vacíos en alcalinidad total de las colecciones de datos de CARINA y GLODAP para el Océano Atlántico.

Además, los valores estimados de A_T derivados del MLR son útiles para el control de calidad de las medidas del sistema del carbonato, ya que pueden ayudar en la identificación de valores atípicos o anómalos (*outliers*). A modo de comparación se diseñó un algoritmo de redes neuronales capaces de realizar predicciones no lineales. El objetivo fue diseñar un enfoque alternativo para llevar a cabo la misma tarea de completar los huecos de alcalinidad total. Ambos métodos generaron resultados internamente consistentes.

1. Introducción

GLODAP y CARINA son colecciones de datos de muestras de botella obtenidas de la síntesis de numerosas campañas oceanográficas (115 y 188 respectivamente). GLODAP ofrece una cobertura global razonable excepto para el Ártico. CARINA se centra en el Atlántico con un énfasis particular en el Atlántico Norte y el Océano Ártico. Ambas colecciones son el resultado de esfuerzos internacionales por sintetizar todos los datos disponibles en productos compilados. El carbono inorgánico total disuelto (C_T), la alcalinidad total (A_T), la concentración total de ion hidrógeno (pH) y la fugacidad del CO_2 (fCO_2) son los cuatro parámetros medidos del sistema del carbonato en agua de mar. Una gran fracción de las muestras en estas bases de datos cuenta con al menos un parámetro del carbono medido, siendo la distribución inicial la mostrada en el panel izquierdo de la Figura 3.1, reproducida de (Velo et al., 2010b). El sistema del carbonato en agua de mar es gobernado por las leyes termodinámicas indicadas en el primer capítulo, y se puede considerar carac-

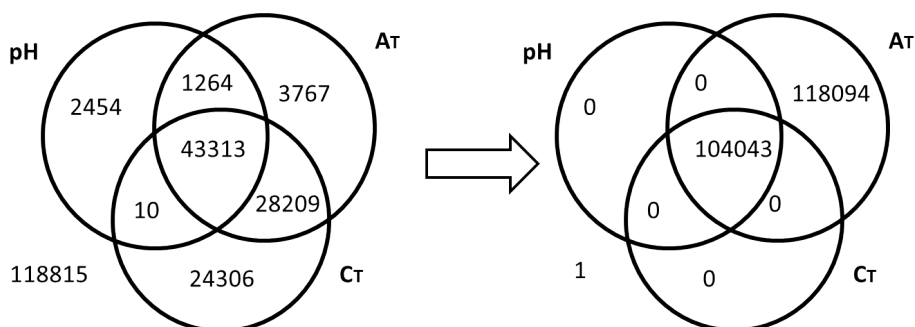


Figura 3.1: Diagrama de Venn indicando la disponibilidad de datos del sistema del carbonato en los datos originales de CARINA/GLODAP empleados y después de añadir los obtenidos mediante los cálculos MLR; los números representan la cantidad de muestras en cada dominio de pH, A_T , C_T o intersección de estos, en la esquina inferior izquierda se muestra la cantidad de datos sin ninguno de los parámetros indicados.

terizado cuando se dispone de datos de al menos dos de los cuatro citados parámetros.

La alcalinidad total es el parámetro del sistema del carbonato que puede ser estimado con más precisión debido a su baja variabilidad en el océano (Wolf-Gladrow et al., 2007; Zeebe and Wolf-Gladrow, 2001). La A_T se supone que está afectada sólo marginalmente por el incremento en las concentraciones de C_T , al menos a la escala decenal que se usa actualmente, debido a que el sistema del carbonato tiende a amortiguar la alcalinidad cuando se disuelve CO_2 en agua de mar actual (Ilyina et al., 2009; Riebesell et al., 2010). Como se ha indicado en el capítulo anterior, la mayor parte de la variabilidad natural de A_T es conocida, y está bien correlacionada con la salinidad, el silicato, e incluso la temperatura (Lee et al., 2006; Millero et al., 1998; Wallace, 1995). Esto permite emplear directamente el conjunto completo de datos de alcalinidad, desde campañas antiguas realizadas en los 70s hasta las más recientes, como si de una climatología se tratase, para los cálculos. No hay, por tanto, necesidad

de dividir el conjunto de datos en series de tiempo o introducir dicho parámetro en las estimaciones, lo que podría resultar en menos datos disponibles en cada partición, o menor precisión en las estimaciones. Por otra parte, es necesario tener en consideración que las bases de datos de partida incluyen medidas anteriores a la década de 1990, y por tanto realizadas sin el empleo de material de referencia certificado (CRMs), medidas, que se ha demostrado, son menos precisas.

Para el trabajo presentado, se ha partido de las colecciones de datos de botella de acceso público de GLODAP para la región atlántica, y de los productos de CARINA para los océanos Atlántico y Austral. Como paso inicial para preparar los datos, se unieron las bases de datos y se realizó un control inicial de calidad y filtrado de valores atípicos. Una vez desarrollado y optimizado el algoritmo, se empleó también para la recuperación de datos de nitrato y fosfato en los casos en los que el dato de botella fuera anómalo o inexistente. Para silicato se empleó una técnica de interpolación multilineal (Velo et al., 2010b) sobre WOA09 (Garcia et al., 2010) en los casos de ausencia de dato original.

2. Metodología

Se propusieron dos técnicas diferentes. En primer lugar un método específico de MLR en ventana móvil 3D, afinado y adaptado a este objetivo (en adelante 3DwMLR), y en segundo lugar un método mediante una red neuronal a medida, también parametrizada y puesta a punto para esta tarea. Ambas alternativas proporcionan aproximaciones diferentes al mismo proble-

ma y por tanto una comparación de ambos resultados puede ser empleada como herramienta para la validación del análisis.

2.1. Método 3DwMLR

El esquema clásico de regresión multilineal permite el modelado de una variable de respuesta y varias predictoras mediante una ecuación lineal. El problema que nos encontramos con el esquema clásico de MLR al aplicarlo al conjunto de datos completo de CARINA y GLODAP es que no toda la variabilidad natural puede ser modelada de este modo, ya que se generarían errores grandes. La relación entre A_T y salinidad y temperatura podría asumirse lineal, pero únicamente cuando los datos se restringen a áreas específicas confinadas en cuencas, y principalmente para aguas superficiales (Lee et al., 2006; Millero et al., 1998).

El objetivo de este trabajo es intentar mejorar los resultados de un MLR clásico combinando las siguientes técnicas:

1. Uso de una ventana móvil 3D para definir la caja de datos alrededor del nodo para el cual intentamos calcular la A_T , de modo que el MLR se realice únicamente con los datos en el interior de la caja.
2. Empleo de un algoritmo de regresión robusta en vez del clásico OLS (*Ordinary Least Squares*).
3. Inclusión de otros parámetros en la regresión en adición a la salinidad y temperatura (Lee et al., 2006; Millero et al., 1998) que también tienen

influencia directa o indirecta en la A_T , y que son: silicato, nitrato, fosfato y oxígeno.

Ventana móvil 3D

Se diseñó un algoritmo para la extracción del conjunto de datos dentro de cada caja para cada muestra. El uso de una ventana móvil ya se planteó para el artículo de interpolación multiparamétrica (Capítulo II: Velo et al., 2010b), y el algoritmo empleado en este trabajo comparte planteamientos con aquel. De hecho, los tamaños de la ventana tanto horizontalmente como en profundidad se parametrizaron acordes al mencionado artículo, aunque se analizaron múltiples combinaciones para mejorar los resultados.

En resumen, se emplearon ventanas móviles de tamaño fijo y variable, aplicando un criterio de mínimo de datos disponibles sobre todas ellas. Las pruebas de tamaño variable comenzaron con una ventana cuadrada que circunscribe un radio equivalente a 2° de latitud, y permitiendo que el algoritmo incrementase el tamaño grado a grado hasta un máximo de 20° de radio, para conseguir al menos 100 muestras dentro de la ventana. En profundidad, se empleó la misma ecuación del artículo de interpolación multiparamétrica (Capítulo II: Velo et al., 2010b), pero incrementando el rango en un 5 % con cada aumento de grado en radio. Las pruebas con tamaños de ventana fijos empleaban un área más amplia de 5° , 10° o incluso 20° de radio en latitud alrededor del nodo. En profundidad se empleó la misma ecuación anterior sin incremento adicional.

Se observó que las parametrizaciones de tamaño variable se ajustaban en mayor medida a las muestras cercanas, pero también resultaban más afectadas por la dispersión de los datos. Las parametrizaciones de ventana fija, sin embargo tenían tendencia a suavizar la variabilidad natural e incluso añadir ruido si la ventana era muy amplia.

Se concluyó que la parametrización más consistente con las observaciones de alcalinidad, era la realizada empleando una ventana “quasi-fija” de 10° en latitud. En realidad se trata de una ventana de tamaño variable que parte de un ancho inicial de 10° de latitud, pero a la que se le exige un mínimo de 100 datos, y para ello se le permite ampliar el tamaño del modo anteriormente descrito. La calificación de “quasi-fija” que se le ha dado se debe a que, de un total de 222136 estimaciones de A_T realizadas por el software, únicamente 5324 (2.4 %) necesitaron una ventana mayor de 10° .

Regresión robusta

Como técnica de regresión multilineal se empleó un algoritmo de regresión robusta (Filzmoser et al., 2007; Hubert, 2008). La regresión robusta combina un algoritmo de regresión clásica por mínimos cuadrados (OLS) con la detección de valores atípicos mediante iteraciones para reponderación de los valores predictores. La técnica también se conoce como mínimos cuadrados con reponderación iterativa (IRLS: *Iteratively Reweighted Least Squares*) (Green, 1984). La reponderación se realiza utilizando una función de ponderación bicuadrada sobre los residuos. De este modo, la alcalinidad inferida es más consistente con las observaciones que cuando se emplea un método OLS.

En la regresión OLS es una tarea crítica mantener los valores de entrada bajo estricto control, ya que el OLS es muy sensible a valores atípicos. Empleando una técnica de regresión robusta, la función de reponderación es capaz de detectar la mayoría de valores atípicos, descartándolos u otorgándoles un impacto mínimo en la regresión. Se empleó una función bicuadrada para la reponderación $(1 - (r/k)^2)^2$ si $r < k$ y 0 en el resto de casos; r representa los residuos escalados por un factor de proporción que utiliza la desviación absoluta respecto a la mediana (MAD: *Median Absolute Deviation*) y k corresponde al valor de corte de 4.685 (Holland and Welsch, 1977). De este modo, los pesos caen rápidamente en cuanto los residuos se desvían significativamente de cero, y los valores son directamente descartados cuando los residuos exceden el valor de corte. La técnica de regresión robusta permanece estadísticamente robusta hasta un 50 % de valores atípicos, mientras que para la regresión OLS, un único valor atípico ya influye en el resultado.

Selección de predictores

Se emplearon los siguientes: presión, temperatura potencial, salinidad, nitrato, fosfato, silicato y oxígeno. El objetivo con la incorporación de todos estos parámetros es incorporar toda la variabilidad posible para obtener una mejor predicción en cada regresión. Los parámetros se seleccionaron por estar relacionados con la alcalinidad y presentar mayor abundancia que los parámetros de carbono en los conjuntos de datos.

Como se indicó en el primer capítulo, la relación entre A_T y ambos, salinidad y silicato es bien conocida. Otros autores han incluido en la parametrización

en aguas superficiales la temperatura potencial (Lee et al., 2006; Millero et al., 1998). La relación de la A_T con nutrientes (nitrato y fosfato) está relacionada con la síntesis y remineralización de la materia orgánica, que añade o elimina estos iones de la disolución y por tanto afecta a la alcalinidad total (Chen and Pytkowicz, 1979; Chen and Millero, 1979). Además, se puede considerar que la suma de A_T , fosfato y las especies del nitrógeno permanece constante durante la remineralización de la materia orgánica (Brewer et al., 1975; Fraga and Álvarez-Salgado, 2005). El oxígeno también influye en la alcalinidad total debido a su uso en los procesos de síntesis y remineralización de la materia orgánica. La relación de la presión con la alcalinidad está relacionada con la formación de carbonato y su disolución a grandes profundidades en el océano. Todos estos procesos han sido tratados extensivamente en la literatura (Dickson, 2010; Millero, 2007; Wolf-Gladrow et al., 2007; Zeebe and Wolf-Gladrow, 2001)

La influencia de los mencionados predictores sobre la regresión también se analizó desde un punto de vista estadístico, para verificar que la selección es adecuada y que todos los parámetros aportan información al modelo. La existencia de colinealidades y correlaciones entre predictores es conocida, pero desde un punto de vista de predicción no se deben eliminar si su presencia mejora los resultados. Por otra parte, el uso de una ventana móvil implica cambios en las condiciones de cada regresión y la importancia relativa de cada predictor en esa ventana concreta. A nivel estadístico se analizó el VIF (*Variance Inflation Factor*) para una regresión multilineal del conjunto de datos completo. Se aplicaron procedimientos similares a los empleados por

Parámetro	VIF(7)	VIF(6)	VIF(5)	VIF(4)	Nº par. usados	ΔAIC	R^2
Nitrato	117.15				7	0.0	0.930
Fosfato	116.26	17.18			6	1539.1	0.929
T. potencial	14.41	10.88	5.54		5	9883.7	0.923
Oxígeno	5.76	4.97	1.29	1.19	4	13249.6	0.921
Salinidad	5.31	4.97	3.57	1.46	3	15282.8	0.919
Silicato	3.67	3.50	1.77	1.59	2	273075.5	0.039
Presión	1.93	1.86	1.75	1.32	1	276923.6	0.002

Tabla III–1: Comprobaciones estadísticas de colinealidad y relevancia de predictores

Juranek et al. (2009) y Kim et al. (2010). En la Tabla III–1, reproducida de (Velo et al., 2013) se muestran en detalle los resultados, que muestran que existe una alta colinealidad cuando se emplea nitrato y fosfato, lo que es ampliamente conocido (Redfield, 1934), pero igualmente es conocido que esta relación presenta diversas fuentes de variabilidad (Anderson and Sarmiento, 1994; Geider and La Roche, 2002). Otro criterio estadístico que se analizó fue el AIC (*Akaike Information Criterion*), también para el modelo completo, y que nos indica una mejor significancia del modelo para valores de AIC más bajos; la Tabla III–1 muestra también estos resultados. Por último, se calculó el estadístico R^2 de la regresión para el conjunto total de datos, que marca el mejor ajuste cuando se emplean los siete parámetros.

Dentro de cada ventana móvil, se comprobó también la distribución de los valores de los predictores alrededor del valor del nodo. Si el valor del nodo está dentro de la nube de datos (una desviación estándar alrededor del promedio), se descartan los valores más allá de tres desviaciones estándar alrededor del nodo. Si la mayoría de valores están alejados del valor del nodo, se descarta por completo ese predictor para ese nodo, con un límite de hasta dos predictores descartados. El objetivo es evitar extrapolaciones con datos fuera de rango que podrían producir valores anómalos. El método asume que las relaciones entre

los predictores y la alcalinidad total son aproximadamente lineales dentro de la ventana móvil, y esto se puede asumir para ventanas pequeñas y predictores en el rango de los valores del nodo.

2.2. Estimación por redes neuronales.

Se planteó inicialmente como método alternativo al MLR que permite ofrecer una alternativa diferente al mismo problema. La red neuronal es capaz de ofrecer una respuesta no lineal, y se entrena, además de con información de masas de agua, con información geográfica, por lo que la fragmentación de datos en ventanas móviles ya no es necesaria. El planteamiento, por tanto, ya es diferente al del MLR. Además, para obtener una solución mediante una red neuronal, en primer lugar necesitamos diseñar y configurar una topología de red que sea adecuada para resolver nuestro problema e implementarla. A continuación necesitamos analizar las alternativas para entrenar la red, y seguidamente entrenarla con datos reales hasta que consideremos que se ajusta lo mejor posible a estos, evitando sobreajustes. Por último emplearemos los predictores de los nodos para obtener unos resultados mediante la red entrenada.

Diseño

Se escogió una red de propagación hacia delante (a nivel topología) con una capa oculta de respuesta log-sigmoidal y otra lineal (Gardner and Dorling, 1998). La ventaja que presenta esta arquitectura es que la capa log-sigmoidal puede asimilar las relaciones no lineales, mientras que la final procesaría las

respuestas lineales (Bishop, 1995; Hagan et al., 1996). En una red neuronal, cada nodo aplica una función de transferencia sobre las entradas para obtener las salidas, empleando pesos que fueron ajustados en el proceso iterativo de aprendizaje. La Figura 3.2, reproducida de (Velo et al., 2013) muestra un esquema de la red empleada.

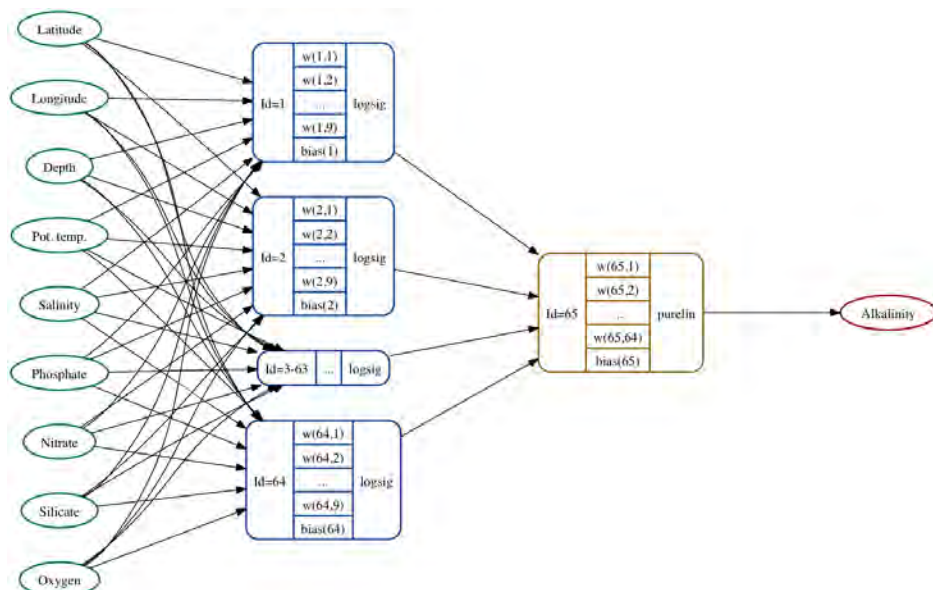


Figura 3.2: Diagrama esquemático de la red neuronal diseñada para la predicción de A_T . Los nodos con Id=1-64 representan la primera capa de respuesta log-sigmoidal, y el nodo con Id=65 la lineal (purelin). Los nodos 3-63 se representan con los puntos en medio de cada caja. Los valores $w(x,y)$ junto con los $bias(x)$ son los pesos establecidos mediante el aprendizaje, y son los que permiten que la red sea útil para predicción.

Los parámetros geográficos y de profundidad se emplearon para informar a la red de la posición de la muestra, y los parámetros físico-químicos para informar de las propiedades de la muestra de agua de mar. Los parámetros fisicoquímicos empleados son los mismos que para el 3DwMLR.

En cuanto al tamaño de la red, no hay un modo de establecer a priori el número de neuronas en la capa oculta (Sarle, 1997), el procedimiento habitual es evaluar la confianza de las predicciones para tamaños diferentes. Se analizaron, por tanto, tamaños desde 2^1 a 2^9 neuronas, obteniéndose el mejor compromiso entre infraestimación y sobreestimación con un tamaño de 64 neuronas en la capa oculta.

Entrenamiento

Las redes neuronales de propagación hacia delante son entrenadas habitualmente con algoritmos de retro-propagación (Hagan et al., 1996; Sarle, 1997). Estas técnicas alimentan la red iterativamente con los valores del conjunto de datos para entrenamiento y propagan hacia atrás los errores de los resultados a través del reajuste de los pesos determinados en la red. Existe una gran variedad de técnicas en la literatura, que se diferencian principalmente en cómo estos pesos son modificados en la retropropagación. En el artículo Velo et al. (2013) se muestra una descripción más detallada de las técnicas analizadas y su respuesta en función del tamaño de la red.

Una técnica que se emplea habitualmente en el entrenamiento, y que se utiliza para mitigar, en parte, los errores, es fragmentar el conjunto de datos que se emplean para el entrenamiento. De este modo, se separan dos pequeños subconjuntos de datos del conjunto principal, en nuestro caso mediante submuestreo aleatorio en cada iteración, y se emplean, uno de ellos para validar el progreso de cada iteración del entrenamiento y el otro para comparación una vez finalizado el entrenamiento. Esta técnica es útil también para ayudar

a detectar el punto óptimo de entrenamiento de la red, ya que, habitualmente, a medida que progresa el entrenamiento los errores que se detectan en el subconjunto de validación disminuyen progresivamente hasta alcanzar un mínimo, pero a partir de ese punto comienzan a incrementarse de nuevo debido a problemas de sobreajuste.

Sobreajuste

Es uno de los problemas más habituales con los sistemas de redes neuronales. La capacidad de una red neuronal de asimilar la información de un modelo y reproducirlo es enorme, llegando a ser capaz de producir residuos nulos entre los datos de entrada y la predicción. Esta capacidad viene determinada por una elección adecuada del modelo, diseño, ajuste y entrenamiento, pero una vez concretados todos estos elementos, la capacidad de ajuste depende principalmente, y es muy sensible, del número de neuronas en las diferentes capas. En los casos de uso en los que se intenta que la red se ajuste lo más posible al modelo esto, generalmente, no supone problema alguno, pero al intentar modelar datos de medida de parámetros reales con sus incertidumbres sí que lo es, ya que pretendemos que la red asimile la información de la variabilidad natural del medio pero que al mismo tiempo no asimile los errores que existen en las medidas de los parámetros de entrada. En la Figura 3.3 se presenta un ejemplo. Si analizamos los ajustes ilustrados por las curvas roja y azul empleando un criterio únicamente basado en los residuos, obtendríamos que el mejor ajuste es el obtenido por la curva roja, cuando es posible que el modelo real se ajuste más a una función como la de la curva azul. Para una

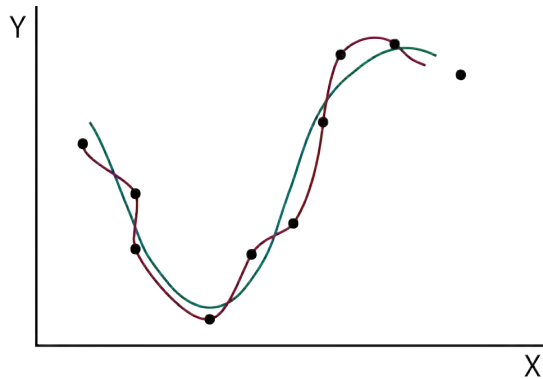


Figura 3.3: Ejemplo de sobreajuste. La tendencia real se representa con la línea azul. Los puntos representarían mediciones del parámetro, con cierto error sobre el valor real. El sobreajuste se representa con la línea roja, que se observa que se aleja de la tendencia real al asimilar los errores de cada punto.

topología de red determinada con unas predicciones como las de la curva azul, un aumento en el número de neuronas provocaría la asimilación de los errores de cada punto, resultando en la curva roja, con un ajuste, sin embargo, casi perfecto.

Dado que no hay modo de conocer, a priori, el número de neuronas adecuado para resolver adecuadamente un determinado problema, la solución pasa por efectuar varios ensayos de prueba-error con diferentes tamaños.

3. Evaluación

Como datos de partida se emplearon las bases de datos de GLODAP (Key et al., 2004; Sabine et al., 2005) y CARINA (Key et al., 2010; Tanhua et al., 2009) combinadas, que inicialmente contaban con un total de 282853 muestras y de las cuales, la A_T estaba presente en 84734 (29.96 %). Se realizaron controles de calidad iniciales evaluando la falta de parámetros (presión o pro-

fundidad, temperatura, salinidad, nutrientes y oxígeno) y valores anómalos (Velo et al., 2013), dando por resultado la eliminación de 60715 muestras, con lo que la base de datos de partida se redujo a 222138 observaciones, con 76553 muestras de A_T . Sobre este conjunto, se aplicaron las metodologías 3DwMLR y neuronal, lo que dio lugar a la generación de 212137 y 212136 estimaciones respectivamente, es decir, todo el conjunto excepto 1 y 2 muestras respectivamente.

Las estimaciones de A_T para las muestras en las que también se dispusiese de dato medido de A_T fueron usadas para evaluar la calidad de ambos métodos de estimación. Se empleó la desviación estándar de los residuos entre dato estimado y dato medido como criterio de calidad, obteniéndose un valor de $5.4 \mu\text{mol}\cdot\text{kg}^{-1}$ para el método 3DwMLR y un valor de $5.2 \mu\text{mol}\cdot\text{kg}^{-1}$ para la red neuronal con entrenamiento mediante el algoritmo Levenberg-Maquardt y 64 neuronas en la capa oculta. Estos resultados indican un buen ajuste entre observaciones y predicciones, ya que los valores son ligeramente inferiores al umbral mínimo para ajuste empleado en el proyecto CARINA para A_T (Tanhua et al., 2010b). Empleando a continuación un intervalo límite de dos veces la desviación estándar anterior, se identificaron 3462 valores (3DwMLR, 4.5 %) y 3526 valores (neuronal, 4.6 %) para los cuales la diferencia entre dato estimado y observado estaba fuera de este límite y que podrían ser etiquetados como cuestionables o incluso reemplazados por el valor estimado.

En las Figuras 3.4 y 3.5 se muestra un resumen de la evaluación de los ajustes obtenidos. La primera columna muestra una comparación de ambos métodos entre sí, y las dos restantes entre cada método y los datos observados.

La escala es $\pm 15 \mu\text{mol}\cdot\text{kg}^{-1}$ y las líneas punteadas verticales de color gris indican los valores -12, -6, 0, 6 y 12 para facilitar la comparación con el límite de $\pm 6 \mu\text{mol}\cdot\text{kg}^{-1}$ de CARINA. En cada caja individual de las figuras de cajas, la altura es proporcional al número de muestras que forman parte de cada intervalo.

3.1. Comparación entre 3DwMLR y red neuronal

El histograma en la parte superior izquierda de la Figura 3.4 indica que existe una buena concordancia entre las técnicas de estimación 3DwMLR y neuronal, con una desviación estándar de $2.36 \mu\text{mol}\cdot\text{kg}^{-1}$ y una media de $0.074 \mu\text{mol}\cdot\text{kg}^{-1}$ para los residuos entre ambas, por lo que el 95 % de los residuos cae dentro del intervalo de $\pm 4.7 \mu\text{mol}\cdot\text{kg}^{-1}$.

Analizando los gráficos de cajas que se muestran en la primera columna de las Figuras 3.4 y 3.5, se observa buena concordancia en general, especialmente analizando el panel relativo al identificador de campaña (*Cruise ID*, Figura 3.4), que indica coherencia, en general, entre ambos métodos para las mismas campañas. Los desajustes en este panel también pueden ser útiles para focalizar el análisis sobre los motivos de los sesgos en esas campañas.

Analizando el resto de paneles de las figuras, se puede observar también que la peor concordancia entre métodos se obtiene para muestras con baja salinidad y altos valores de oxígeno y que, identificando su procedencia, se asimilarían con muestras superficiales (profundidades menores de 200 m) cercanas a costa y afectadas por aportes continentales, como la Corriente del Labrador y zonas más antárticas del Océano Austral. La variabilidad, en estos casos

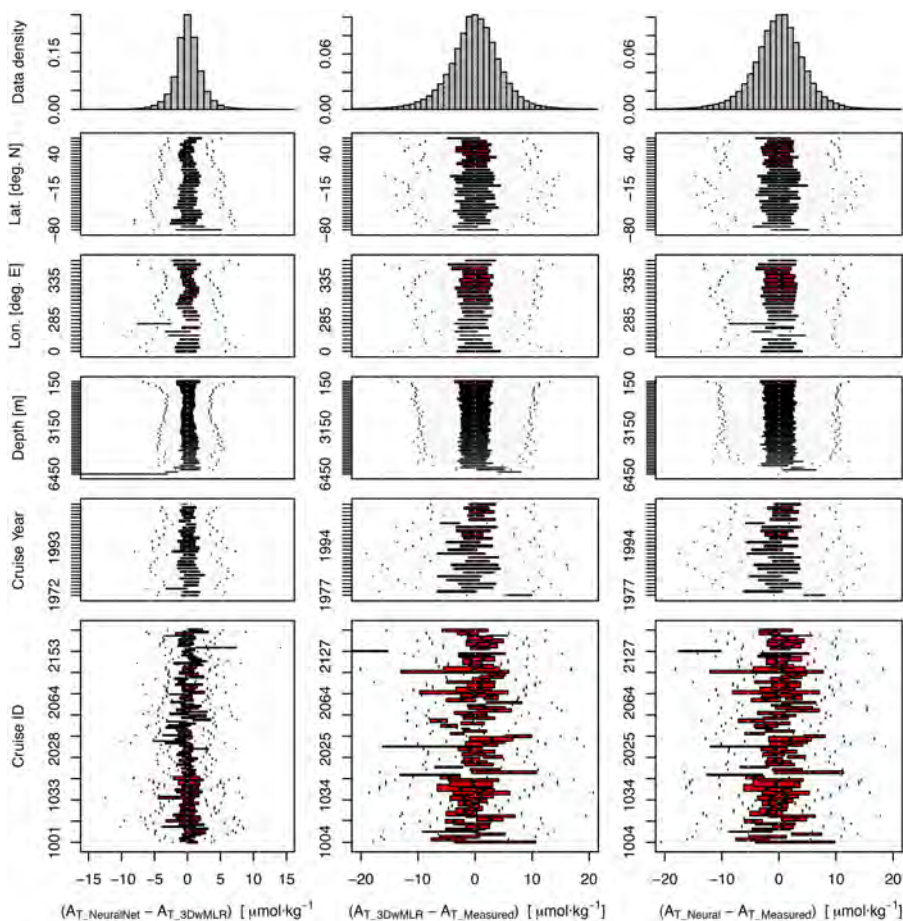


Figura 3.4: Histogramas y diagramas de cajas de los residuos de la comparación de los métodos de estimación de A_T . De izquierda a derecha: estimación Neuronal menos estimación a 3DwMLR, estimación 3DwMLR menos observaciones y estimación Neuronal menos observaciones. De arriba abajo, la primera fila es el histograma de densidad de datos, y los siguientes diagramas muestran los residuos de A_T frente al parámetro indicado en el eje izquierdo: latitud, longitud, profundidad, año de campaña y por último un identificador numérico de cada campaña.

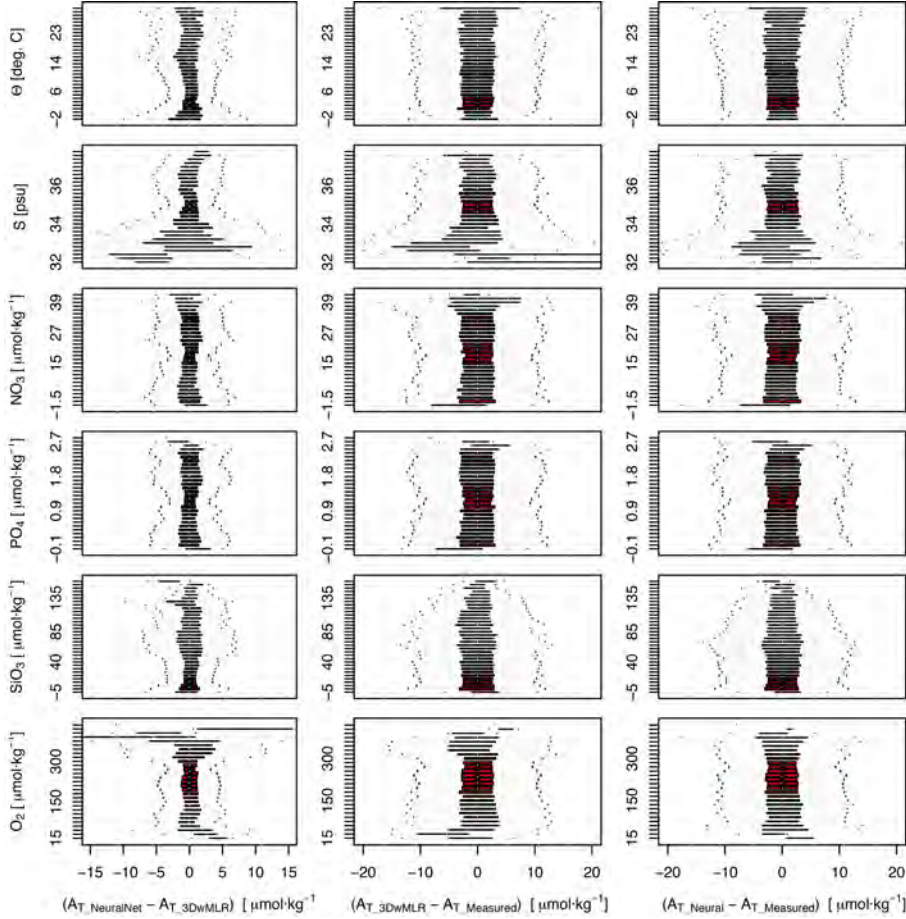


Figura 3.5: Diagramas de cajas de los residuos de la comparación de los métodos de estimación de A_T . De izquierda a derecha: estimación Neuronal menos estimación a 3DwMLR, estimación 3DwMLR menos observaciones y estimación Neuronal menos observaciones. De arriba abajo, los siguientes diagramas muestran los residuos de A_T frente al parámetro indicado en el eje izquierdo: temperatura potencia, salinidad, nitrato, fosfato, silicato y oxígeno

es debida a la mezcla de aguas en escalas cortas de tiempo (Rintoul, 2011; Wootton et al., 2008), que no es explicada por los métodos de estimación.

3.2. Comparación entre métodos y observaciones

El análisis se presenta en los paneles de la columna central para el método 3DwMLR y en la columna derecha para la red neuronal de las Figuras 3.4 y 3.5. También muestran una buena concordancia en general, y se puede apreciar en el panel relativo al año de campaña (*Cruise Year*, Figura 3.4), cómo la dispersión de los residuos es menor para las campañas más recientes, lo cual es consistente con la mejora en las técnicas analíticas y la instrumentación.

Analizando visualmente los residuos obtenidos frente a los parámetros físicos y bioquímicos del agua de mar (Figura 3.5), se observan patrones similares para ambas metodologías. El método Neuronal parece tener comportamientos más consistentes para la mayoría de parámetros, pero las diferencias no son significativas. Asimismo, al igual que se indicó anteriormente para la comparación entre métodos, en la comparación con observaciones también se observa que la mayor dispersión está relacionada con muestras de valores bajos de salinidad; sin embargo, en este caso las muestras con mayores concentraciones de oxígeno muestran una buena correlación entre métodos y observaciones. Tal y como se comentaba en el apartado anterior, este hecho podría estar motivado por la variabilidad de corta escala de tiempo en la mezcla de aguas que los métodos no son capaces de interpretar, obteniendo tendencias diferentes.

Por otro lado, analizando el panel relativo al identificador de campaña (*Cruise ID*, Figura 3.4), se observa también un desajuste significativo para algunas

Carina ID	Expocode	3DwMLR (desajuste)	Red Neuronal (desajuste)	Ajuste propuesto
9	06MT19920509	-7.8±9.0	-7.6±8.6	-7.7
28	06MT20021013	-7.2±19.8	-3.8±15.8	Flag bad
51	29CS19771007	7.6±3.9	6.0±2.6	6.8
57	29GD19860904	-6.9±4.0	-5.1±4.5	-6.0
158	67SL19881117	-19.0±4.8	-14.1±6.0	-16.6

Tabla III-2: Desajustes promedio entre los métodos 3DwMLR y Neuronal frente a los datos de A_T observados. Las unidades de los desajustes son $\mu\text{mol}\cdot\text{kg}^{-1}$. Se muestran únicamente los valores con un desajuste medio superior a $\pm 6 \mu\text{mol}\cdot\text{kg}^{-1}$.

campañas y que, sin embargo no se percibe en la comparación entre métodos, lo que implica que ambos métodos coinciden en el diagnóstico de dicho desajuste. En la Tabla III-2 se muestran las campañas para las que se han obtenido desajustes promedio superiores a $\pm 6 \mu\text{mol}\cdot\text{kg}^{-1}$ y que se corresponden con campañas con pocas o ninguna muestra de profundidades superiores a 1500 m, y que por lo tanto no han sido consideradas en el control de calidad (QC2) mediante *crossovers* de CARINA (Tanhua et al., 2010b, 2009; Velo et al., 2009) como candidatas para ajuste. Como resultados de este trabajo, se proponen los ajustes indicados en la Tabla 5 para las campañas indicadas en dicha tabla.

Capítulo IV

Capítulo IV – Estudio de tasas de almacenamiento de C_{ant} en el Atlántico Suroeste

1. Introducción

La cuenca Oeste del Océano Atlántico Sur, desde los 10°N a los 55°S y desde la costa hasta la dorsal atlántica está caracterizada por procesos dinámicos complejos que implican varias células de circulación en los niveles superiores e intermedios (Mémery et al., 2000). Es una región con grandes incertidumbres en cuanto al almacenamiento de carbono antropogénico (C_{ant}) ya que presenta concentraciones moderadas de C_{ant} en aguas profundas ($\approx 10 \mu\text{mol}\cdot\text{kg}^{-1}$) pero un gran volumen de agua. El análisis de datos de las tres últimas décadas indica una tasa de almacenamiento de C_{ant} de $0.92 \pm 0.13 \text{ mol}\cdot\text{m}^{-2}\cdot\text{y}^{-1}$, es decir, un 13-35 % superiores que las estimaciones previas en este área. Las pequeñas pero significativas concentraciones de C_{ant} en el gran volumen de la relativamente bien ventilada Agua de Fondo Antártica (AABW), podrían ser la causa subyacente de esta tasa de almacenamiento, de hecho, la contribución significativa en términos de C_{ant} de esta AABW ventilada que entra en la región del Océano Atlántico Sur se calculó en $0.055 \pm 0.02 \text{ Pg}\cdot\text{C}\cdot\text{y}^{-1}$ o $0.20 \text{ mol}\cdot\text{m}^{-2}\cdot\text{y}^{-1}$. En vez de basarla en una tendencia anual, la evolución del inventario específico

de C_{ant} (en $\text{mol}\cdot\text{m}^{-2}$) se computa más consistentemente como una función de la perturbación de la $x\text{CO}_2$ atmosférica en ppm ($0.64 \text{ mol}\cdot\text{m}^{-2}\cdot\text{ppm}^{-1}$). Esta metodología permite proyecciones mejoradas de las tasas de almacenamiento de C_{ant} sobre períodos largos.

2. Metodología

Para realizar el presente estudio, se partió de un grupo de 20 campañas oceanográficas de extenso recorrido, que se indican en la Tabla IV-1, y que están disponibles en el área de estudio en las bases de datos de GLODAP (Key et al., 2004; Sabine et al., 2005) y CARINA (Hoppema et al., 2009; Key et al., 2010; Tanhua et al., 2009). El grupo de campañas empleado cubre un abanico temporal de 33 años (1975-2005), y fue agrupado en seis conjuntos centrados cada uno en un año de referencia para mejorar la cobertura espacial en cada intervalo temporal. Los años de referencia se indican en la primera columna de la Tabla IV-1. Se consideraron únicamente los datos de profundidad superior a 100 m para evitar la alta variabilidad de las medidas superficiales. Por otro lado, y dado que en las campañas más antiguas las mediciones tanto de A_T como de C_T fueron realizadas por técnicas potenciométricas y sin el uso de material de referencia (CRMs), se aplicaron las correcciones que se indican a continuación.

Las campañas TTO-TAS (transectos 1 y 2) y SAVE (transectos 1 y 2) fueron corregidas siguiendo las indicaciones de Gruber et al. (1996); la campaña 3230CITHER2 (Sección A17, 1993) fue corregida aplicando una corrección de $-8 \mu\text{mol}\cdot\text{kg}^{-1}$ sobre A_T reportada en Ríos et al. (2005). Por otro lado,

Año de referencia	Año de campaña	Id de campaña	CARINA expocode / GLODAP code	Sección WOCE o nombre original
1972	1972	43	GEOSECS 1-2	GEOSECS
1983	1982-3	46	TTOTAS 1-3	TTO-TAS
1988	1987-8	48	318MSAVE 1-5	SAVE
1993	1991	13	06MT15/3	A09
1993	1991	22	OACES91 1-2	A16S
1993	1993	14	06MT22/5	A10
1993	1993	23	OACES93	A16N
1993	1994	12	06MT19941012	
1993	1994	21	316N142	A15
1993	1994	24	3230CITHER2 1-2	A17
1993	1995	95	35LU19950909	
1997	1996	84	33LK19960415	
1997	1997	25	316N151 3	A20
1997	1999	106	35TH19990712	
2003	2001	61	29HE20010305	FICARAM
2003	2002	62	29HE20020304	FICARAM
2003	2003	68	316N20030922	
2003	2003	86	33RO20030604	
2003	2003	113	49NZ20031106	Beagle
2003	2005	87	33RO20050111	

Tabla IV-1: Campañas de la cuenca oeste del Atlántico Sur empleadas en este estudio. Las campañas fueron agrupadas en seis conjuntos centrados en los años de referencia indicados en la primera columna.

los datos de la campaña GEOSECS presentaban gran dispersión, por lo que se aplicó un método MLR (Capítulo III: Velo et al., 2013) sobre el conjunto completo de CARINA y GLODAP para obtener un valor estimado de A_T . Este valor se empleó como criterio para detectar datos potencialmente anómalos (separados más de $12 \mu\text{mol}\cdot\text{kg}^{-1}$) que fueron descartados.

Se empleó también la base de datos en cuadrícula de WOA09 (Antonov et al., 2010; Locarnini et al., 2010) para caracterizar la región de estudio en capas de densidad, siguiendo los intervalos definidos Mémerly et al. (2000) y con los que se computó un espesor medio para cada capa. Estos valores se presentan en la Tabla IV-2. Para realizar este cálculo se realizó una interpolación de los datos del perfil de densidad en columna de WOA para cada nodo geográfico,

Masa de agua	Espesor de capa (m)	Intervalos de anomalía de densidad ($\text{kg}\cdot\text{m}^{-3}$)
SACW	168±32	100 m a $\sigma_\theta < 26.5$ (300 m)
SAMW	306±25	$26.5 < \sigma_\theta < 27.1$ (600 m)
AAIW	412±14	$27.1 < \sigma_\theta < 27.4$ (1100 m)
uNADW-uCDW	1594±68	$\sigma_\theta > 27.4$ y $\sigma_3 < 41.47$ (3000 m)
INADW-ICDW	532±166	$\sigma_3 > 41.47$ y $\sigma_4 < 45.90$ (4000 m)
AABW	902±311	$\sigma_4 > 45.90$ (hasta fondo)

Tabla IV-2: Intervalos de densidad (σ_x) para las seis capas que definen las principales masas de agua de la región. El espesor de las capas se indica en metros (m). El intervalo de confianza definido por la desviación estándar. Entre paréntesis se indican las profundidades nominales de las isopícnas

obteniendo así los rangos de profundidad que permiten calcular los espesores en cada nodo geográfico. Por último se realizó un promediado para obtener el espesor medio de cada capa en la región de estudio.

Las estimaciones de carbono antropogénico fueron realizadas aplicando el método φC_T^0 (Vázquez-Rodríguez et al., 2009a) sobre los datos de botella de las campañas seleccionadas para el estudio. Una vez obtenido el dato de C_{ant} para cada muestra de botella, se aplicó la metodología de Pérez et al. (2010) para el cómputo de los inventarios específicos de C_{ant} que se definen según la ecuación (4.1):

$$Inv_c^{C_{\text{ant}}} = \sum_{l=1}^6 \rho^{l,c} \cdot C_{\text{ant}}^{l,c} \cdot Th^{l,c} \quad (4.1)$$

En la que $Inv_c^{C_{\text{ant}}}$ representa el inventario específico para el conjunto de campañas c ; $\rho^{l,c}$ representa la densidad promedio del conjunto de campañas l en la capa l , y que se obtiene para las posiciones de las campañas mediante la climatología de WOA09; $C_{\text{ant}}^{l,c}$ representa la concentración de C_{ant} promedio para el conjunto de campañas c en la capa l ; y $Th^{l,c}$ representa el espesor

promedio de la capa l para la campaña c obtenido a partir de los datos climatológicos de WOA09 por el procedimiento indicado anteriormente.

Para evitar los problemas derivados de la baja cobertura espacial y la dispersión en años de las campañas analizadas, se realizó un análisis MLR para estimar un valor de corrección de concentración de C_{ant} para cada capa. Como parámetros predictores se emplearon la salinidad, temperatura potencial, AOU, nutrientes y el propio año de la campaña (Pérez et al., 2010). Los coeficientes obtenidos del MLR fueron multiplicados por la anomalía promedio del valor frente a WOA09, es decir, la diferencia entre la media de los valores de estos predictores para la capa y la media de los valores de las “estaciones” (perfil en columna de agua para cada nodo geográfico) de WOA09 equivalentes para esos parámetros. De este modo, se obtuvo un término de corrección $\Delta C_{ant}^{l,c}$ que se añadió al C_{ant} estimado por botellas $C_{ant}^{l,c} = \Delta C_{ant}^{l,c} + C_{ant,botella}^{l,c}$

A continuación se analizó la evolución temporal obtenida de los promedios de C_{ant} para cada una de las seis capas a lo largo del periodo temporal que abarcan las campañas. La Figura 4.1, reproducida de (Ríos et al., 2012) muestra los resultados de este análisis y presenta los valores de concentración de C_{ant} de cada capa frente (a) al año de la campaña, y (b) frente a la ΔxCO_2 (asumiendo la xCO_2 de referencia de 278.5 ppm). En ambos paneles se representa con línea punteada y en cuadrados rojos el inventario de C_{ant} para la cuenca. En el panel (a) también se representa con una línea roja fina la evolución del inventario que se obtendría teniendo en cuenta la tasa de almacenamiento calculada del panel (b) (línea roja punteada) que es de 0.64 ± 0.13 mol/m²/ppm, y la ΔxCO_2 para cada año basada en los datos de Mauna Loa

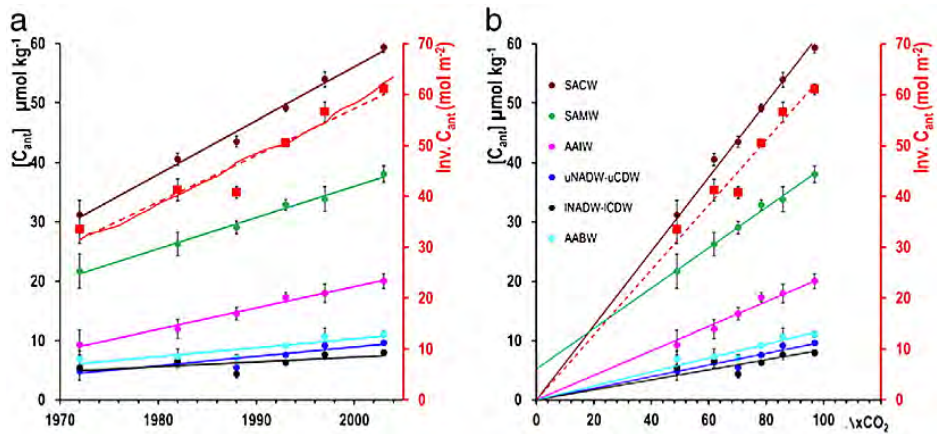


Figura 4.1: Evolución de los promedios de concentración de C_{ant} en $\mu\text{mol}\cdot\text{kg}^{-1}$ (puntos) e inventarios específicos de C_{ant} en $\text{mol}\cdot\text{C}\cdot\text{m}^{-2}$ (cuadrados) frente a año (a) y frente a exceso de $x\text{CO}_2$ atmosférico ($\Delta x\text{CO}_2$ en ppm) (b), para la capa de cada masa de agua (leyenda en gráfica b). Las barras de error representan $\pm 2\sigma/N^{0.5}$ (donde σ es la desviación estándar y N el número de datos usados). El incremento de concentración de C_{ant} (línea continua, ejes izquierdos) en (a) $\mu\text{mol}\cdot\text{kg}^{-1}\cdot\text{y}^{-1}$ y (b) $\mu\text{mol}\cdot\text{kg}^{-1}\cdot\text{ppm}^{-1}$, y las tasas de almacenamiento de C_{ant} (línea punteada, ejes derechos) en (a) $\text{mol}\cdot\text{m}^{-2}\cdot\text{y}^{-1}$ y (b) en $\text{mol}\cdot\text{m}^{-2}\cdot\text{ppm}^{-1}$ se reproducen en la Tabla IV-3. Las tasas de almacenamiento de C_{ant} en $\text{mol}\cdot\text{m}^{-2}\cdot\text{y}^{-1}$ (línea delgada) fueron calculadas a partir de las tasas de almacenamiento de C_{ant} de $0.64 \text{ mol}\cdot\text{m}^{-2}\cdot\text{ppm}^{-1}$ para cada año según el $\Delta x\text{CO}_2$ correspondiente de Mauna Loa, y representadas como inventario multiplicadas por 0.0169 (eje derecho).

(Hawaii). Se puede observar como la línea roja continua del panel (a) sigue el patrón general del inventario y la tasa lineal (en función de año), pero se separa un poco en los extremos, haciendo notar que la tendencia real es de tipo exponencial. En ambos paneles y en la Tabla IV-3 se puede observar también como los almacenamientos de C_{ant} decrecen a medida que las capas se vuelven más profundas, a excepción de la AABW, con valores más altos que los de las capas inmediatamente superiores (INADW-ICDW y uNADW-uCDW)

Masa de agua	Tasa ($\mu\text{mol}\cdot\text{kg}^{-1}\cdot\text{y}^{-1}$)	R ²	p- level	Tasa ($\mu\text{mol}\cdot\text{kg}^{-1}\cdot\text{ppm}^{-1}$)	R ²	p- level
SACW	0.90±0.04	0.991	0.002	0.627±0.006	0.999	<0.001
SAMW	0.53±0.02	0.991	0.002	0.34±0.015	0.991	<0.001
AAIW	0.36±0.02	0.985	0.004	0.207±0.004	0.998	<0.001
uNADW-uCDW	0.16±0.04	0.820	0.052	0.098±0.004	0.990	<0.001
INADW-ICDW	0.08±0.04	0.450	0.250	0.085±0.006	0.970	<0.001
AABW	0.15±0.04	0.800	0.061	0.118±0.004	0.990	<0.001
	mol·m⁻²·y⁻¹			mol·m⁻²·ppm⁻¹		
Tasa para la columna completa	0.92±0.13	0.930	0.019	0.64±0.13	0.997	<0.001

Tabla IV-3: Tasa de incremento de concentración de C_{ant} en $\mu\text{mol}\cdot\text{kg}^{-1}\cdot\text{y}^{-1}$ y $\mu\text{mol}\cdot\text{kg}^{-1}\cdot\text{ppm}^{-1}$ en las diferentes capas de las masas de agua. Los errores representan $\pm 2\sigma/N^{0.5}$. En la parte inferior se representan las tasas de almacenamiento de C_{ant} en $\text{mol}\cdot\text{m}^{-2}\cdot\text{y}^{-1}$ y $\text{mol}\cdot\text{m}^{-2}\cdot\text{ppm}^{-1}$ para la columna de agua completa de la cuenca oeste del Atlántico Sur

3. Evaluación

Las tasas de almacenamiento anual observadas para la cuenca Oeste del Atlántico Sur en este trabajo, de $0.92\pm 0.13 \text{ mol}\cdot\text{m}^{-2}\cdot\text{y}^{-1}$, son superiores a la de estimaciones previas, como se puede observar en la Tabla IV-4, reproducida de (Ríos et al., 2012).

Un análisis detallado permite ver que las principales diferencias con los estudios anteriores se deben a la estimación del C_{ant} para la capa de AABW que entra en el Atlántico Sur. Esta capa, de bajo contenido en C_{ant} , representa sin embargo un gran volumen de agua por lo que la cantidad de C_{ant} almacenada es significativa. Este C_{ant} es difícilmente detectable con métodos de tipo MLR si el intervalo de tiempo entre muestras es pequeño (inferior a 2 décadas).

Por otro lado, el cómputo de la tasa de almacenamiento frente al incremento de xCO_2 atmosférica, estimado en este artículo en $0.64\pm 0.13 \text{ mol}\cdot\text{m}^{-2}\cdot\text{ppm}^{-1}$ (Ríos et al., 2012) permite obtener estimaciones a más largo plazo, debido a la correspondencia del incremento de CO_2 en la atmósfera con una función

Autor	Método	Tasa de almacenamiento ($\text{mol}\cdot\text{m}^{-2}\cdot\text{y}^{-1}$)
Murata et al. (2008)	Isopycnal	0.80
Peng and Wanninkhof (2010)	MLR (eMLR)	0.74 ± 0.3 (0.35) ^a
Wanninkhof et al. (2010)	eMLRdens	0.60
Este estudio	Backcalculation	0.92 ± 0.13

^a El valor entre paréntesis refiere a la tasa de almacenamiento para el eMLR

Tabla IV-4: Comparación de almacenamientos de C_{ant} ($\text{mol}\cdot\text{m}^{-2}\cdot\text{y}^{-1}$) para la cuenca oeste del Océano Atlántico Sur (10°N a 55°S)

exponencial. Esta tasa de almacenamiento es de utilidad para proyecciones de inventarios futuros de C_{ant} en diferentes escenarios de $x\text{CO}_2$ atmosférica debido a la linealidad existente entre ambos términos.

El análisis realizado en este artículo (Ríos et al., 2012) para la cuenca oeste del Océano Atlántico Sur destaca la necesidad de continuar con programas de campañas en secciones repetidas, ya que este tipo de estudios ayudan a mejorar la detección de las variaciones de C_{ant} en aguas profundas y de fondo, que son de relevancia por representar un gran volumen del océano.

Conclusiones

Conclusiones

Los dos artículos que conforman el primer capítulo de la tesis (Velo et al., 2010a, 2009) presentan una evaluación de la calidad de los datos obtenidos tras los procesos de síntesis de A_T y pH, y que se centra, por tanto, en analizar la calidad de dichos datos mediante el apoyo de métodos estadísticos de regresión. Mediante el uso de estas técnicas y confinando en lo posible los datos en regiones para intentar limitar su variabilidad y así poder asumir cierta linealidad, es factible analizar la correlación entre parámetros. Dado que para cada parámetro se ha realizado un control de calidad independiente, este análisis permite detectar problemas de inter-consistencia entre parámetros al aplicar los ajustes propuestos, o la ausencia de estos. Se procedió dividiendo los datos en capas de densidad, según las principales masas de agua de la región de trabajo, y realizando un análisis gráfico de los residuos que se obtienen como diferencia entre los valores del respectivo parámetro (A_T y pH) y los obtenidos mediante un método de regresión multilineal. Este método se aplicó sobre los datos originales de partida, y también sobre los datos del producto final (obtenidos tras la síntesis: QC2 y aplicación de ajustes), permitiendo, de este modo, analizar si el resultado de la síntesis minimiza los residuos.

Se concluye por tanto que en ambos casos (A_T y pH), los residuos obtenidos son menores empleando la base de datos corregida en comparación con la original sin corregir. La mayoría, pero no todas las campañas del producto, obtienen un valor de mediana para los residuos dentro del límite de ± 6

$\mu\text{mol}\cdot\text{kg}^{-1}$ para A_T y de ± 0.005 para pH, lo que entra dentro del error esperable en el análisis de laboratorio de dichos parámetros. Es necesario tener en cuenta que este error de los residuos está influenciado por los errores de la medición de la alcalinidad y pH, pero también por los de medida de los parámetros predictores y de la propia aplicación del método MLR que no es capaz de explicar la totalidad de la variabilidad natural observada.

Las tareas a realizar dentro del proyecto CARINA se dividieron y repartieron entre los numerosos participantes en dicho proyecto. Por tanto, las conclusiones ofrecidas en los artículos, se centran en el trabajo realizado de supervisión y análisis de los resultados de los procedimientos de control de calidad empleados en el proyecto CARINA para A_T y pH, y, pese a describirlos en la sección de métodos, no abordan conclusiones sobre los procedimientos empleados o sobre el propio proyecto. Estas conclusiones se incluyen en otros artículos expresamente dedicados a ello, y en los cuales tanto el autor de esta Tesis como su director han participado asumiendo diversas tareas. Los trabajos que complementan este capítulo son, por tanto, Tanhua et al. (2010b, 2009) de procedimientos y general del proyecto, Key et al. (2010) del producto de datos, y Van Heuven et al. (2009) de cálculos del sistema del carbonato en lenguaje Matlab.

De un modo más general debemos resaltar los avances que el proyecto CARINA ofrece globalmente a la comunidad oceanográfica. CARINA toma como punto de partida las lecciones aprendidas tras el proyecto GLODAP, pero con una indiscutible ventaja adicional: el desarrollo de la tecnología entre mediados de los 90 (GLODAP) y segunda mitad de los 2000 es tal que elimina el

cuello de botella que podía suponer el acceso individual a las herramientas necesarias para el tratamiento de grandes cantidades de datos y, además, ofrece nuevas herramientas mejoradas, tanto a nivel de capacidades de procesamiento como de aplicaciones. Esto permite que la práctica totalidad de integrantes del proyecto CARINA cuenten con la capacidad tecnológica para analizar y procesar los datos.

En cuanto al enfoque o modo de proceder, es necesario destacar la necesidad de desarrollar cada fase del proyecto de modo que pueda ser reproducida de un modo lo más sencillo posible, y que permita, en todo momento, trazar los pasos realizados.

El desarrollo de las fases de un proyecto de síntesis de datos a partir de datos de campañas desarrolla sus fases de un modo secuencial, tal y como se describe en la Figura 1 de Tanhua et al. (2010b). Ello implica que cada fase parte y depende de los resultados de la anterior y, por ello, que habitualmente sea necesario repetir por completo fases posteriores ante cambios o correcciones en las anteriores. Esto sucede habitualmente ante la incorporación de nuevas campañas, o modificaciones en los datos originales de éstas en las fases finales del proyecto, y por el análisis de los resultados de los controles de calidad de cada fase, que también ayudan a detectar errores anteriores. Esto da lugar a un proceso quasi-iterativo en el desarrollo del proyecto con la repetición continua de fases posteriores.

Es por ello recomendable implementar los procedimientos a realizar en cada fase de modo que:

- Sea lo más sencillo posible repetirlos.

- Sea posible obtener un registro o traza de los procedimientos aplicados a cada dato individual, y decisiones tomadas, para facilitar la detección, localización y corrección de problemas.
- A medida que se avance en el proceso, se vayan incorporando procedimientos que permitan, en la medida de lo posible, detectar los errores más comunes automáticamente.

La solución más común para llevarlo a cabo es la implementación de los procesos vía rutinas o procedimientos de software que no sólo permiten automatizar las tareas en lo posible, sino también analizar a posteriori qué cálculos fueron realizados. Ambos puntos son igualmente importantes. Sin embargo, dada la heterogeneidad de los datos y la gran cantidad de muestras, estaciones y campañas a procesar, es igualmente necesario realizar una muy cuidadosa supervisión manual de las distintas fases y sus resultados debido a que, pese a la incorporación progresiva de mecanismos de control, a lo largo del desarrollo del proyecto se detectan continuamente factores que los procesos automáticos no tienen en cuenta, y que producen o pueden producir un sesgo importante en los resultados.

El siguiente paso natural tras la liberación del producto de síntesis del proyecto CARINA que se compone de muestras de estaciones en posiciones y profundidades discretas, es la realización de un mallado que permita obtener un producto en una cuadrícula homogénea, que es la tarea que se describe en el segundo capítulo. En el momento de la realización de dicha tarea, los antecedentes eran la interpolación del producto del proyecto anterior, GLODAP, generado mediante supercomputadoras de la década de los 90, y algoritmos

como DIVA o métodos de interpolación objetiva. Dada la gran dispersión de los datos en algunas zonas (Océano Austral y Atlántico Sur), se planteó que el empleo de distancias basadas en parámetros físico-químicos conservativos en vez de espaciales, junto con la incorporación de una base de datos con mucha mayor resolución en esos parámetros como base para el mallado, podría obtener resultados mejores que cualquier interpolación puramente espacial.

Procediendo de este modo, el mallado en zonas dispersas seguiría los patrones termohalinos de esa zona al incorporar la climatología de WOA para la propagación de los valores. Tras el desarrollo de los procedimientos, se analizó el resultado del mallado de Oxígeno, comparándolo contra el propio de WOA. Por último se generó una base de datos en cuadrícula para el C_{ant} en el Océano Atlántico aplicando el método φC_T^0 centrado al 1994 (Vázquez-Rodríguez et al., 2009b) sobre los datos de síntesis y mallando el resultado.

La conclusión principal de este capítulo es que el método WMP obtiene mejores resultados que un método espacial tradicional. Mediante el empleo, como base de datos auxiliar, de WOA (Antonov et al., 2006; Garcia et al., 2006a, 2006b; Locarnini et al., 2006) en su versión de 2005, que ha sido elaborada con una mayor resolución de datos biogeoquímicos, el método WMP cuenta con más información para la tarea de mallado que cualquier otra alternativa exclusivamente espacial.

En cuanto a los inventarios de C_{ant} (referido al año 1994) obtenidos, los valores oscilan entre 55 y 58 Pg-C en función de la técnica de estimación empleada (respectivamente φC_T^0 o TrOCA). Es necesario concluir también, que la técnica de interpolación empleada no tiene apenas efecto sobre las

estimaciones de inventario, y ello es debido a efectos de compensación entre diferentes dominios. Sin embargo, analizando las regiones en detalle, existen algunas pequeñas diferencias entre los resultados obtenidos por ambos métodos de interpolación. La interpolación WMP aparenta funcionar mejor que la puramente espacial en zonas con menor densidad de datos, como se puede observar en las figuras, con especial hincapié sobre y bajo el límite de la isolínea de 5°C. El método puramente espacial tiende a producir valores más bajos (más altos) de C_{ant} para las aguas bajo (sobre) la isoterma de 5°C en comparación con el método multiparamétrico WMP.

El tercer capítulo parte a su vez de las lecciones aprendidas en el desarrollo de los capítulos anteriores, y propone el desarrollo de un método que permita estimar valores de uno de los parámetros clave en el sistema del carbono, la A_T , con el objetivo de que los valores generados puedan servir tanto para incorporar un nivel adicional de control de calidad sobre el parámetro (evolución sobre el capítulo I) como para ampliar la base de datos de muestras de A_T , lo que mejora la cantidad de datos disponibles, entre otros, para su empleo en estimaciones de C_{ant} (evolución sobre la aplicación del capítulo II). Los procedimientos consistieron en someter al producto de síntesis de Carina a un proceso de análisis de interconsistencia entre A_T y varios parámetros predictores, y mediante dos aproximaciones diferentes; una basada en métodos MLR en los que se trata de conseguir la mayor linealidad posible para predecir datos; y otra mediante redes neuronales con el mismo propósito de la predicción. Se puede concluir que los desarrollos realizados obtienen un error estimado en la predicción bastante aceptable para el momento actual.

De una base de datos inicial en el producto de síntesis de CARINA con 72796 muestras útiles para cálculos de carbono, es decir, con al menos dos parámetros del sistema del carbono medidos, el algoritmo desarrollado permite extender la base de datos hasta las 104043 muestras, lo que implica un 42.9 % de incremento en la abundancia de datos útiles.

Por otra parte, los resultados obtenidos mediante ambos métodos propuestos (neural y MLR) obtienen resultados aceptables, ya que el error de predicción es inferior a $5 \mu\text{mol}\cdot\text{kg}^{-1}$, que en el momento actual se considera un error aceptable para la medición de la alcalinidad y parejo al establecido en el proyecto CARINA ($6 \mu\text{mol}\cdot\text{kg}^{-1}$).

Es necesario destacar, sin embargo, que las dos aproximaciones para resolver el problema son muy diferentes, ofreciendo cada una de ellas ventajas e inconvenientes que se deben considerar:

Desde un punto de vista estrictamente de rendimiento de ambos métodos en cuanto a la capacidad de modelar la variabilidad natural, el método neuronal ofrece mejores resultados. Los resultados que se muestran en la Figura 3 del artículo de estimación de A_T (Velo et al., 2013) indican claramente que la parametrización propuesta con 64 neuronas en la capa oculta y el empleo del algoritmo de Levenberg-Marquadt para el aprendizaje, obtiene mejores estimaciones, en general, que el método 3DwMLR, con menores errores promedio entre predicción y datos originales.

Desde un punto de vista de escalabilidad de la metodología, sin embargo, el método 3DwMLR es la mejor opción, ya que ofrece un buen ajuste a los datos medidos y presenta la ventaja de que la metodología puede ser reajustada o

adaptada. En el método 3DwMLR la totalidad de pasos para cada estimación individual puede ser depurada y así mejorada. La red neuronal representa una alternativa totalmente opuesta, ya que es cierto que una red entrenada puede ser re-entrenada, pero el conjunto de pesos obtenido no es transparente, hasta el punto de que la combinación de una parametrización de red neuronal en conjunto con los resultados de su entrenamiento pueden ser considerados como una caja negra que funciona para un determinado conjunto de datos.

La ventaja del empleo simultáneo de ambas alternativas en este capítulo es que, de este modo, sus resultados pueden ser fácilmente comparados entre sí, y con ello determinar su rendimiento y la concordancia obtenida.

Dado que la presente tesis doctoral trata sobre la optimización de métodos para el análisis de la acidificación en los océanos, no estaría completa sin incluir un análisis de acidificación en el que se haga uso de métodos y enfoques aquí propuestos. El último capítulo se presenta, por tanto, como la participación en un estudio en el que se aplican algunas de las técnicas indicadas anteriormente en el análisis de las tasas de almacenamiento de C_{ant} para la cuenca oeste del Atlántico Sur, focalizado sobre el rol del Agua de Fondo Antártica (AABW)

El estudio emplea un subconjunto de 20 campañas extraídas de las bases de datos de síntesis de CARINA y GLODAP. Dado que para las campañas más antiguas tanto A_T como C_T fueron medidos potencialmente y sin el empleo de CRMs (su aparición y extensión de uso fue posterior), la aplicación de un método MLR para analizar, en lo posible, la calidad de los datos de A_T era importante. Del estudio se puede destacar que las tasas de almacenamiento de C_{ant} observadas para la cuenca oeste del Océano Atlántico Sur, de 0.92 ± 0.13

$\text{mol}\cdot\text{m}^{-2}\cdot\text{y}^{-1}$, son más altas que las de estimaciones previas. Analizando en detalle las diferencias, se observa que el origen de las discrepancias se puede asociar con el AABW que entra en el Atlántico Sur. Aunque la tasa de almacenamiento de C_{ant} que se estima partiendo de métodos de retro-cálculo es pequeña, el gran volumen de esta masa de agua provoca que sea relevante tanto a efectos de inventario como de tasas de almacenamiento. La tasa asociada al AABW se estima en $0.20 \text{ mol}\cdot\text{m}^{-2}\cdot\text{y}^{-1}$, que representa un 22 % del total determinado en este estudio. Las desviaciones entre las tasas obtenidas por el método φC_T^0 y las de métodos como “*isopycnal surfaces*”, MLR y eMLR son debidas a la penetración de aguas profundas y de fondo con baja concentración de C_{ant} . Estas bajas concentraciones son casi indetectables, sobre todo para métodos MLR cuando se emplean datos de periodos inferiores a dos décadas.

Por otro lado, dado que el incremento exponencial de CO_2 atmosférico afecta a la linealidad de las estimaciones a largo plazo, en este estudio se ha calculado también la tasa de almacenamiento de C_{ant} como una función lineal sobre el exceso de xCO_2 atmosférico sobre un estado preindustrial ($0.64\pm0.13 \text{ mol}\cdot\text{m}^{-2}\cdot\text{ppm}^{-1}$), asumiendo el concepto de “estado estacionario transitorio” para el C_{ant} (Steinfeldt et al., 2009; Tanhua et al., 2007). Así computada, esta tasa presenta la ventaja de la linealidad sobre el incremento exponencial de CO_2 atmosférico, lo que la hace útil para proyecciones de inventarios en diferentes escenarios de xCO_2 atmosférico.

Los análisis realizados aquí sobre el Atlántico Sur, así como los realizados en el resto de capítulos, en particular el primero y el segundo, ponen de manifiesto

la necesidad de continuar programas de secciones repetidas con mediciones de alta calidad de parámetros del sistema del carbono, ya que son fundamentales para la correcta estimación de la evolución de las tasas de almacenamiento e inventarios de carbono antropogénico.

Por otro lado la evolución actual en los proyectos de integración de datos de observación, es hacia minimizar el intervalo de tiempo desde la observación de los datos hasta su liberación a la comunidad científica. La adquisición de datos oceanográficos, por las propias características del medio, tiene un coste muy elevado y, en ocasiones, desde la propia medición hasta la liberación de los datos pasan años, o incluso nunca se llegan a liberar. El enfoque actual pretende analizar y revisar todos los pasos y procesos que se llevan a cabo entre ambos hitos con el objetivo de optimizar y automatizar en lo posible los procesos. Trabajos como el presentado en esta tesis pretenden avanzar en esa dirección.

Trabajo actual y futuro

La experiencia adquirida en la realización de los trabajos de control de calidad primario y secundario sobre los parámetros A_T en el Océano Atlántico y pH a nivel global, nos ha permitido formar parte del equipo de trabajo del proyecto GLODAPv2. GLODAPv2 es la segunda versión del *Global Ocean Data Analysis Project* y pretende unificar la práctica totalidad de bases de datos existentes de carbono e incorporar la mayor cantidad de campañas individuales posibles, para obtener un producto de síntesis con los controles de calidad realizados, y que en el momento de su publicación será la mayor base de datos global de parámetros relevantes al carbono en el océano. Actualmente consiste en cerca de 700 campañas y más de 45000 estaciones, cubriendo el período de 1972 a 2013. GLODAPv2 incorpora los productos de GLODAPv1 y CARINA y ha logrado incorporar también la base de datos PACIFICA, de reciente creación y con 306 campañas en el Océano Pacífico. A mayores incorpora cerca de 100 nuevas campañas.

Las tareas realizadas y a realizar en el ámbito de este proyecto son, por un lado la unificación de escalas de datos de pH para toda la colección global de campañas; la colaboración en la realización de las tareas de control de calidad primario y secundario, con el análisis de los resultados obtenidos mediante la aplicación de versiones actualizadas de las rutinas *cnaX* de *crossover* creadas por S. Van Heuven (Tanhua et al., 2010b) y por último la participación en el grupo de interpolación para el suministro de un método de interpolación

que ofrezca una versión alternativa a DIVA, que permita también ofrecer un producto de GLODAPv2 en cuadrícula homogénea y así comparar resultados.

En el ámbito del proyecto GLODAPv2, un método de interpolación como el WMP descrito en el artículo (Velo et al., 2010b) que constituye el segundo capítulo de la tesis, presenta grandes ventajas frente a una interpolación puramente espacial, ya que constituye un escenario similar al de CARINA para el cual fue diseñado. Sin embargo, el software desarrollado presentaba limitaciones que evitaban poder aplicarlo directamente a la interpolación a nivel global. Se ha intentado mejorar el método para hacer esto posible. Las debilidades que se han intentado solucionar se indican a continuación:

1.- Capacidad para interpolar varias cuencas de modo autónomo y sin información adicional. Las rutinas de software originales permitían interpolar sin problema alguno el Océano Atlántico, y sería posible aplicarlas individualmente a cada cuenca para interpolar la totalidad de océanos, pero se ha preferido mejorar el algoritmo para que evitase cruzar fronteras de tierra automáticamente. Para ello se ha empleado la malla equiespaciada de ETOPO2V2 como referencia batimétrica y topográfica para discriminar océano y tierra, y un algoritmo de tipo “*raytracing*” (trazado de líneas) alrededor de cada nodo a interpolar para evitar datos de estaciones tras un límite terrestre. Ha sido necesario por otro lado, optimizar el algoritmo al objeto de poder discriminar islas de pequeño tamaño.

2.- Capacidad para interpolar a altas latitudes, especialmente en los polos. Las rutinas del algoritmo de interpolación WMP extraen los datos alrededor de un nodo mediante una ventana ortoédrica, empleando límites en latitud,

longitud y profundidad. El motivo es únicamente un compromiso de precisión frente a rendimiento, ya que la aplicación de rutinas de cálculo de distancias incrementaría en más de un orden de magnitud el tiempo total de interpolado, lo que es contrario al objetivo. En el desarrollo de las rutinas de interpolación del capítulo II se analizó la diferencia entre el uso de uno u otro algoritmo en la interpolación del Océano Atlántico y se observó que las mejoras con el cálculo de distancias esféricas eran mínimas frente al incremento en tiempo de cómputo. Dado que GLODAPv2 incluye datos del Ártico hasta los 90°N de latitud, la asunción original obligaba a replantear el algoritmo. Esto motivó el desarrollo de un nuevo algoritmo, esta vez sí basado en distancias, concretamente distancias por círculo máximo, que permite extraer los datos de estaciones dentro de un radio determinado alrededor de cada nodo. La forma de la ventana pasa, por tanto, de ser ortoédrica a cilíndrica. Se realizaron optimizaciones mediante la identificación a priori de estaciones para permitir la extracción en un tiempo contenido, con lo que finalmente se mantuvo el tiempo de cálculo próximo al tiempo original.

3.- Distribución de valores alrededor del nodo. Otro de los problemas que se puso especialmente de manifiesto al realizar un mallado global, fue la carencia de una distribución homogénea de valores alrededor de los nodos de determinadas regiones. Habitualmente este problema se minimiza ampliando el tamaño de la ventana de extracción, que amplían los orígenes de datos y generalmente mejora la distribución. Sin embargo, esta solución no es ideal ya que genera problemas de suavizado excesivo en algunas áreas. Se procedió modificando el algoritmo de extracción de datos de ventana, dividiendo el

círculo alrededor del nodo a interpolar en cuatro cuadrantes y exigiendo un mínimo de datos en al menos tres de estos. El mínimo se definió en un 10 % de la cantidad de valores exigidos para interpolar cada nodo, habitualmente 100.

4.- Cómputo de errores de interpolación. Todavía en desarrollo, se están evaluando alternativas para su estimación. El método implementado actualmente y en evaluación, prácticamente un esquema quasi-montecarlo, consiste en aplicar n veces (actualmente $n=100$) una perturbación sobre los valores de los nodos a interpolar. La perturbación se aplica únicamente sobre los datos del parámetro a interpolar y consiste en una distribución normal y aleatoria de errores con una desviación estándar prefijada y similar a la esperable en la precisión de dicho parámetro. La técnica propuesta limita el cómputo de errores a los propios del parámetro a interpolar y no tiene en cuenta los posibles errores por la precisión en la localización de las muestras o de los parámetros empleados en las distancias multiparamétricas. La ventaja de este esquema se obtiene en tiempo de computación, ya que permite aplicar las perturbaciones sobre las distancias paramétricas de cada nodo ya computadas. De este modo, el cómputo del error supone un incremento de tiempo de cálculo para el global de la interpolación de un 2-5 %, lo que lo hace asumible para la potencia de cálculo de que disponemos.

Actualmente, el algoritmo presenta una mejora considerable frente a alternativas espaciales, ofreciendo interpolaciones correctas para estaciones muy separadas, como las secciones latitudinales del Océano Pacífico, pero permitiendo mantener los detalles y sin caer en suavizados excesivos.

Bibliografía

Bibliografía

- Anderson, L.A., Sarmiento, J.L., 1994. Redfield ratios of remineralization determined by nutrient data analysis. *Glob. Biogeochem Cycles* 8, 65–80, DOI: 10.1029/93GB03318
- Antonov, J.I., Locarnini, R.A., Boyer, T.P., Mishonov, A.V., Garcia, H.E., Levitus, S., 2006. *World Ocean Atlas 2005, Volume 2: Salinity*. S. Levitus. NOAA Atlas NESDIS 62 US Gov. Print. Off. Wash. DC 182
- Antonov, J.I., Seidov, D., Boyer, T.P., Locarnini, R.A., Mishonov, A.V., Garcia, H.E., Baranova, O.K., Zweng, M.M., Johnson, D.R., 2010. *World Ocean Atlas 2009, Volume 2: Salinity*. NOAA Atlas NESDIS 69, 184
- Bates, R., Vijh, A., 1973. Determination of pH: Theory and practice. *J. Electrochem. Soc.* 120, 263C–263C
- Bennett, A.F., 1992. *Inverse methods in physical oceanography*. Cambridge University Press
- Berelson, W.M., Balch, W.M., Najjar, R., Feely, R.A., Sabine, C., Lee, K., 2007. Relating estimates of CaCO_3 production, export, and dissolution in the water column to measurements of CaCO_3 rain into sediment traps and dissolution on the sea floor: A revised global carbonate budget: A REVISED GLOBAL CARBONATE BUDGET. *Glob. Biogeochem. Cycles* 21, n/a–n/a, DOI: 10.1029/2006GB002803
- Bishop, C.M., 1995. *Neural networks for pattern recognition*. Clarendon Press, Oxford
- Bretherton, F.P., Davis, R.E., Fandry, C.B., 1976. A technique for objective analysis and design of oceanographic experiments applied to MODE-73. *Deep-Sea Res.* 23, 559–582
- Brewer, P.G., Wong, G.T.F., Bacon, M.P., Spencer, D.W., 1975. An oceanic calcium problem? *Earth Planet. Sci. Lett.* 26, 81–87, DOI: 10.1016/0012-821X(75)90179-X
- Broecker, W.S., 1974. “NO”, a conservative water-mass tracer. *Earth Planet. Sci. Lett.* 23, 100–107, DOI: 10.1016/0012-821X(74)90036-3
- Broecker, W.S., Peng, T.-H., 1982. *Tracers in the Sea*. Lamont-Doherty Geological Observatory, Columbia University
- Caldeira, K., Wickett, M.E., 2003. Oceanography: Anthropogenic carbon and ocean pH. *Nature* 425, 365–365, DOI: 10.1038/425365a
- Canadell, J.G., Le Quére, C., Raupach, M.R., Field, C.B., Buitenhuis, E.T., Ciais, P., Conway, T.J., Gillett, N.P., Houghton, R.A., Marland, G., 2007. Contributions to accelerating atmospheric CO_2 growth from economic activity, carbon intensity, and efficiency of natural sinks. *Proc. Natl. Acad. Sci.* 104, 18866

- Chen, C.-T.A., Pytkowicz, R.M., 1979. On the total CO₂-titration alkalinity-oxygen system in the Pacific Ocean. *Nature* 281, 362–365, DOI: 10.1038/281362a0
- Chen, G.-T., Millero, F.J., 1979. Gradual increase of oceanic CO₂. *Nature* 277, 205–206, DOI: 10.1038/277205a0
- Clayton, T., Byrne, R.H., 1993. Calibration of m-cresol purple on the total hydrogen ion concentration scale and its application to CO₂-system characteristics in seawater. *Deep-Sea Res.* 40, 211–2129
- Dickson, A., 1981. An exact definition of total alkalinity and a procedure for the estimation of alkalinity and total inorganic carbon from titration data. *Deep-Sea Res.* 28, 609–623
- Dickson, A., 1984. pH scales and proton-transfer reactions in saline media such as sea water. *Geochim Cosmochim Acta* 48, 2299–2308
- Dickson, A., Millero, F., 1987. A comparison of the equilibrium constants for the dissociation of carbonic acid in seawater media. *Deep-Sea Res.* 34, 1733–1743, DOI: 10.1016/0198-0149(87)90021-5
- Dickson, A., Riley, J., 1979. The estimation of acid dissociation constants in seawater media from potentiometric titrations with strong base. I. The ionic product of water (K_w). *Mar. Chem.* 7, 89–99
- Dickson, A.G., 1990a. Standard potential of the reaction: $\text{HSO}_4^- \rightleftharpoons \text{H}^+ + \text{SO}_4^{2-}$, and the standard acidity constant of the ion HSO_4^- in synthetic sea water from 273.15 to 318.15 K. *J. Chem. Thermodyn.* 22, 113–127.
- Dickson, A.G., 1990b. Thermodynamics of the dissociation of boric acid in synthetic seawater from 273.15 to 318.15 K. *Deep Sea Res. Part Oceanogr. Res. Pap.* 37, 755–766.
- Dickson, A.G., 1993. The measurement of sea water pH. *Mar. Chem.* 44, 131–142
- Dickson, A.G., 2010. The carbon dioxide system in seawater: equilibrium chemistry and measurements, in: *Guide to Best Practices for Ocean Acidification Research and Data Reporting*. Publications Office of the European Union, Luxembourg, pp. 17–52
- Dickson, A.G., Sabine, C.L., Christian, J.R., 2007. Guide to best practices for ocean CO₂ measurements [WWW Document]. URL http://cdiac.ornl.gov/oceans/Handbook_2007.html (accessed 8.22.13)
- Emery, W.J., Thomson, R.E., 2001. *Data analysis methods in physical oceanography*. Elsevier Science
- Feely, R.A., Sabine, C.L., Lee, K., Millero, F.J., Lamb, M.F., Greeley, D., Bullister, J.L., Key, R.M., Peng, T.H., Kozyr, A., Ono, T., Wong, C.S., 2002. In situ calcium carbonate dissolution in the Pacific Ocean. *Glob. Biogeochem. Cycles* 16, 1144
- Filzmoser, P., Serneels, S., Maronna, R., Van Espen, P.J., 2007. Robust multivariate methods in chemometrics. *Compr. Chemom. Chem. Biochem. Data Anal.* Elsevier Amst. 681–722

- Fraga, F., Álvarez-Salgado, X.A., 2005. On the variation of alkalinity during phytoplankton photosynthesis. *Cienc. Mar.* 627–639
- Freeland, H.J., Gould, W.J., 1976. Objective analysis of meso-scale ocean circulation features. *Deep-Sea Res.* 23, 915–923
- Gandin, L.S., Hardin, R., 1965. Objective analysis of meteorological fields. Israel program for scientific translations Jerusalem
- Garcia, H.E., Locarnini, R.A., Boyer, T.P., Antonov, J.I., Levitus, S., 2006a. World Ocean Atlas 2005, Volume 4: Nutrients (phosphate, nitrate, silicate). S. Levitus. NOAA Atlas NESDIS 64 US Gov. Print. Off. Wash. DC 396
- Garcia, H.E., Locarnini, R.A., Boyer, T.P., Antonov, J.I., Levitus, S., 2006b. World Ocean Atlas 2005, Volume 3: Dissolved Oxygen, Apparent Oxygen Utilization, and Oxygen Saturation. S. Levitus. NOAA Atlas NESDIS 63 US Gov. Print. Off. Wash. DC 342
- Garcia, H.E., Locarnini, R.A., Boyer, T.P., Antonov, J.I., Zweng, M.M., Baranova, O.K., Johnson, D.R., 2010. World Ocean Atlas 2009, Volume 4: Nutrients (phosphate, nitrate, and silicate). Levitus Ed NOAA Atlas NESDIS 71 US Gov. Print. Off. Wash. DC 398
- Gardner, M., Dorling, S., 1998. Artificial neural networks (the multilayer perceptron) - a review of applications in the atmospheric sciences. *Atmos. Environ.* 32, 2627–2636, DOI: 10.1016/S1352-2310(97)00447-0
- Geider, R., La Roche, J., 2002. Redfield revisited: variability of C:N:P in marine microalgae and its biochemical basis. *Eur. J. Phycol.* 37, 1–17, DOI: 10.1017/S0967026201003456
- Green, P.J., 1984. Iteratively Reweighted Least Squares for Maximum Likelihood Estimation, and some Robust and Resistant Alternatives. *J. R. Stat. Soc. Ser. B Methodol.* 46, 149–192
- Gruber, N., Sarmiento, J.L., Stocker, T.F., 1996. An improved method for detecting anthropogenic CO₂ in the oceans. *Glob. Biogeochem. Cycles* 10, 809–837, DOI: 10.1029/96GB01608
- Hagan, M.T., Demuth, H.B., Beale, M.H., 1996. Neural network design. PWS Publishing, Boston, MA
- Hansson, I., 1973. A new set of pH scales and standard buffers for sea water. *Deep-Sea Res.* 20, 471–491
- Holland, P.W., Welsch, R.E., 1977. Robust regression using iteratively reweighted least-squares. *Commun. Stat.-Theory Methods* 6, 813–827
- Hoppema, M., Velo, A., van Heuven, S., Tanhua, T., Key, R.M., Lin, X., Bakker, D.C.E., Pérez, F.F., Ríos, A.F., Lo Monaco, C., Sabine, C.L., Álvarez, M., Bellerby, R.G.J., 2009. Consistency of cruise data of the CARINA database in the Atlantic sector of the Southern Ocean. *Earth Syst Sci Data* 1, 63–75

- Hubert, M., 2008. High-Breakdown Robust Multivariate Methods. *Stat. Sci.* 23, 92–119, DOI: 10.1214/088342307000000087
- Ilyina, T., Zeebe, R.E., Maier-Reimer, E., Heinze, C., 2009. Early detection of ocean acidification effects on marine calcification. *Glob. Biogeochem. Cycles* 23, 11 PP., DOI: 200910.1029/2008GB003278
- Intergovernmental Oceanographic Commission of UNESCO, 2013. Ocean Data Standards, Vol.3: Recommendation for a Quality Flag Scheme for the Exchange of Oceanographic and Marine Meteorological Data. IOC Manuals and Guides, 54, Vol: 3, Paris
- Jalilkee, J.B., Hamilton, D.R., 1977. Objective analysis and classification of oceanographic data. *Tellus* 29
- Joos, F., Plattner, G.K., Stocker, T.F., Marchal, O., Schmittner, A., 1999. Global warming and marine carbon cycle feedbacks on future atmospheric CO₂. *Science* 284, 464
- Juranek, L.W., Feely, R.A., Peterson, W.T., Alin, S.R., Hales, B., Lee, K., Sabine, C.L., Peterson, J., 2009. A novel method for determination of aragonite saturation state on the continental shelf of central Oregon using multiparameter relationships with hydrographic data. *Geophys. Res. Lett.* 36, 6 PP., DOI: 200910.1029/2009GL040778
- Jutterström, S., Anderson, L.G., Bates, N.R., Bellerby, R., Johannessen, T., Jones, E.P., Key, R.M., Lin, X., Olsen, A., Omar, A.M., 2010. Arctic Ocean data in CARINA. *Earth Syst Sci Data* 2, 71–78
- Key, R.M., Kozyr, A., Sabine, C.L., Lee, K., Wanninkhof, R., Bullister, J.L., Feely, R.A., Millero, F.J., Mordy, C., Peng, T.H., 2004. A global ocean carbon climatology: Results from Global Data Analysis Project (GLODAP). *Glob. Biogeochem. Cycles* 18, GB4031, DOI: 10.1029/2004GB002247
- Key, R.M., Tanhua, T., Olsen, A., Hoppema, M., Jutterström, S., Schirnick, C., van Heuven, S., Kozyr, A., Lin, X., Velo, A., Wallace, D.W.R., Mintrop, L., 2010. The CARINA data synthesis project: Introduction and overview. *Earth Syst Sci Data* 2, 105–121, DOI: 10.5194/essd-2-105-2010
- Kim, T.-W., Lee, K., Feely, R.A., Sabine, C.L., Chen, C.-T.A., Jeong, H.J., Kim, K.Y., 2010. Prediction of Sea of Japan (East Sea) acidification over the past 40 years using a multiparameter regression model. *Glob. Biogeochem. Cycles* 24, 14 PP., DOI: 201010.1029/2009GB003637
- Kutner, M.H., Nachtsheim, C.J., Neter, J., Li, W., 2004. Applied linear statistical models. McGraw-Hill IRWIN Chic
- Le Quéré, C., Rodenbeck, C., Buitenhuis, E.T., Conway, T.J., Langenfelds, R., Gomez, A., Labuschagne, C., Ramonet, M., Nakazawa, T., Metzl, N., 2007. Saturation of the Southern Ocean CO₂ sink due to recent climate change. *Science* 316, 1735

- Lee, K., Choi, S.D., Park, G.H., Wanninkhof, R., Peng, T.H., Key, R.M., Sabine, C.L., Feely, R.A., Bullister, J.L., Millero, F.J., 2003. An updated anthropogenic CO₂ inventory in the Atlantic Ocean. *Glob. Biogeochem. Cycles* 17, 1116, DOI: 10.1029/2003GB002067
- Lee, K., Tong, L.T., Millero, F.J., Sabine, C.L., Dickson, A.G., Goyet, C., Park, G.H., Wanninkhof, R., Feely, R.A., Key, R.M., 2006. Global relationships of total alkalinity with salinity and temperature in surface waters of the world's oceans. *Geophys. Res. Lett.* 33, L19605, DOI: 10.1029/2006GL027207
- Lewis, E., Wallace, D., Allison, L., 1998. Program developed for CO₂ system calculations. ORNL/CDIAC-105, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee
- Lo Monaco, C., Álvarez, M., Key, R.M., Lin, X., Tanhua, T., Tilbrook, B., Bakker, D.C.E., van Heuven, S., Hoppema, M., Metzl, N., Ríos, A., Sabine, C.L., Velo, A., 2010. Assessing the internal consistency of the CARINA database in the Indian sector of the Southern Ocean. *Earth Syst Sci Data* 2, 51–70
- Locarnini, R.A., Mishonov, A.V., Antonov, J.I., Boyer, T.P., Garcia, H.E., Baranova, O.K., Zweng, M.M., Johnson, D.R., 2010. World Ocean Atlas 2009, Volume 1: Temperature. NOAA Atlas NESDIS 68, 184
- Locarnini, R.A., Mishonov, A.V., Antonov, J.I., Boyer, T.P., Garcia, H.E., Levitus, S., 2006. World Ocean Atlas 2005, Volume 1: Temperature. S. Levitus. NOAA Atlas NESDIS 61 US Gov. Print. Off. Wash. DC 182
- Mehrbach, C., Culbertson, C., Hawley, J., Pytkowicz, R., 1973. Measurement of the apparent dissociation constants of carbonic acid in seawater at atmospheric pressure. *Limnol. Oceanogr.* 897–907
- Mémery, L., Arhan, M., Álvarez-Salgado, X.A., Messias, M.J., Mercier, H., Castro, C.G., Ríos, A.F., 2000. The water masses along the western boundary of the south and equatorial Atlantic. *Prog. Oceanogr.* 47, 69–98, DOI: 10.1016/S0079-6611(00)00032-X
- Millero, F., Zhang, J., Fiol, S., Sotolongo, S., Roy, R., Lee, K., Mane, S., 1993. The use of buffers to measure the pH of seawater. *Mar. Chem.* 44, 143–152
- Millero, F.J., 2007. The Marine Inorganic Carbon Cycle. *Chem. Rev.* 107, 308–341, DOI: 10.1021/cr0503557
- Millero, F.J., Lee, K., Roche, M., 1998. Distribution of alkalinity in the surface waters of the major oceans. *Mar. Chem.* 60, 111–130, DOI: 10.1016/S0304-4203(97)00084-4
- Mintrop, L., Pérez, F.F., González Dávila, M., Santana-Casiano, J.M., Körtzinger, A., 2000. Alkalinity determination by potentiometry: Intercalibration using three different methods. *Cienc. Mar.* 26, 23–37
- Morris, A., Riley, J., 1966. The bromide/chlorinity and sulphate/chlorinity ratio in sea water. *Deep-Sea Res.* 13, 699–705

- Murata, A., Kumamoto, Y., Sasaki, K., Watanabe, S., Fukasawa, M., 2008. Decadal increases of anthropogenic CO₂ in the subtropical South Atlantic Ocean along 30 S. *J Geophys Res* 113, C06007, DOI: 10.1029/2007JC004424
- NASA/GISS, 2014. GISS Surface Temperature Analysis [WWW Document]. GISS Surf. Temp. Anal. -Comb. Land-Surf. Air Sea-Surf. Water Temp. Anom. Land-Ocean Temp. Index LOTI
- Olafsson, J., Olafsdottir, S.R., Benoit-Cattin, A., Danielsen, M., Takahashi, T., 2009. North Atlantic Ocean acidification from time series measurements. *IOP Conf. Ser. Earth Environ. Sci.* 6, 462005, DOI: 10.1088/1755-1307/6/46/462005
- Olsen, A., 2009. Nordic Seas total alkalinity data in CARINA. *Earth Syst Sci Data* 2, 79–97
- Orr, J., Fabry, V., Aumont, O., Bopp, L., Doney, S., Feely, R., Gnanadesikan, A., Gruber, N., Ishida, A., Joos, F., 2005. Anthropogenic ocean acidification over the twenty-first century and its impact on calcifying organisms. *Nature* 437, 681–686
- Peng, T.-H., Wanninkhof, R., 2010. Increase in anthropogenic CO₂ in the Atlantic Ocean in the last two decades. *Deep Sea Res. Part Oceanogr. Res. Pap.* 57, 755–770, DOI: 10.1016/j.dsr.2010.03.008
- Pérez, F.F., Fraga, F., 1987. The pH measurements in seawater on the NBS scale. *Mar. Chem.* 21, 315–327, DOI: 10.1016/0304-4203(87)90054-5
- Pérez, F.F., Mouriño, C., Fraga, F., Ríos, A.F., 1993. Displacement of water masses and remineralization rates off the Iberian Peninsula by nutrient anomalies. *J. Mar. Res.* 51, 869–892, DOI: 10.1357/0022240933223891
- Pérez, F.F., Vázquez-Rodríguez, M., Mercier, H., Velo, A., Lherminier, P., Ríos, A.F., 2010. Trends of anthropogenic CO₂ storage in North Atlantic water masses. *Biogeosciences* 7, 1789–1807, DOI: 10.5194/bg-7-1789-2010
- Raven, J., Caldeira, K., Elderfield, H., Hoegh-Guldberg, O., Liss, P., Riebesell, U., Shepherd, J., Turley, C., Watson, A., 2005. Ocean acidification due to increasing atmospheric carbon dioxide. *R. Soc. Lond. Document No.* 12/05
- Redfield, A.C., 1934. On the proportions of organic derivatives in sea water and their relation to the composition of plankton. *James Johnstone Meml. Vol.* 176, 92
- Riebesell, U., Fabry, V.J., Hansson, L., Gattuso, J.P., 2010. Guide to best practices for ocean acidification research and data reporting. Publications Office of the European Union, Luxembourg
- Riley, J., 1965. The occurrence of anomalously high fluoride concentrations in the North Atlantic. *Deep-Sea Res.* 12, 219–220
- Rintoul, S.R., 2011. The Southern Ocean in the Earth System. *Sci. Dipl. Antarct. Sci. Gov. Int. Spaces* 175–188
- Ríos, A.F., Fraga, F., Pérez, F.F., 1989. Estimation of coefficients for the calculation of “NO”, “PO” and “CO”, starting from the elemental composition of natural phytoplankton. *Sci Mar* 53, 779–784

- Ríos, A.F., Johnson, K.M., Avarez-Salgado, X.A., Arlen, L., Billant, A., Bingler, L.S., Branellec, P., Castro, C.G., Chipman, D.W., Roson, G., Wallace, D.W.R., 2005. Carbon Dioxide, Hydrographic, and Chemical Data Obtained During the R/V Maurice Ewing Cruise in the Atlantic Ocean (WOCE Section A17, 4 January - 21 March 1994)
- Ríos, A.F., Velo, A., Pardo, P.C., Hoppema, M., Pérez, F.F., 2012. An update of anthropogenic CO₂ storage rates in the western South Atlantic basin and the role of Antarctic Bottom Water. *J. Mar. Syst.* 94, 197–203, DOI: 10.1016/j.jmarsys.2011.11.023
- Sabine, C., Key, R., Kozyr, A., Feely, R., Wanninkhof, R., Millero, F., Peng, T., Bullister, J., Lee, K., 2005. Global Ocean data Analysis Project (GLODAP): Results and data, NDP-083, 110 pp. Carbon Dioxide Inf Anal Cent Oak Ridge Natl Lab Oak Ridge Tenn
- Sabine, C.L., Feely, R.A., Gruber, N., Key, R.M., Lee, K., Bullister, J.L., Wanninkhof, R., Wong, C.S., Wallace, D.W.R., Tilbrook, B., Millero, F.J., Peng, T.-H., Kozyr, A., Ono, T., Ríos, A.F., 2004. The Oceanic Sink for Anthropogenic CO₂. *Science* 305, 367–371, DOI: 10.1126/science.1097403
- Sabine, C.L., Hoppema, M., Key, R.M., Tilbrook, B., van Heuven, S., Lo Monaco, C., Metzl, N., Ishii, M., Murata, A., Musielewicz, S., 2009. Assessing the internal consistency of the CARINA data base in the Pacific sector of the Southern Ocean. *Earth Syst Sci Data Discuss* 2, 555–578
- Sarle, W.S., 1994. Neural networks and statistical models. *Proc. Ninet. Annu. SAS Users Group Int. Conf. SAS Inst. Inc* 1538–1550
- Sarle, W.S., 1997. Neural Network FAQ. *comp.ai.neural-nets*
- Sørensen, S., 1909. Enzymstudien II. Über die Messung und die Bedeutung der Wasserstoffionen-Konzentration bei enzymatischen Prozessen. *Biochem Z* 21, 131–199
- Steinfeldt, R., Rhein, M., Bullister, J.L., Tanhua, T., 2009. Inventory changes in anthropogenic carbon from 1997–2003 in the Atlantic Ocean between 20°S and 65°N. *Glob. Biogeochem. Cycles* 23, GB3010, DOI: 10.1029/2008GB003311
- Takahashi, T., Williams, R., Bos, D., 1982. Carbonate chemistry. *GEOSECS Pac. Exped.* 3, 1973–1974
- Tanhua, T., 2010. Matlab Toolbox to Perform Secondary Quality Control (2nd QC) on Hydrographic Data. ORNL/CDIAC-158. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee. DOI: 10.3334/CDIAC/otg.CDIAC_158
- Tanhua, T., Körtzinger, A., Friis, K., Waugh, D.W., Wallace, D.W., 2007. An estimate of anthropogenic CO₂ inventory from decadal changes in oceanic carbon content. *Proc. Natl. Acad. Sci.* 104, 3037, DOI: 10.1073/pnas.0606574104
- Tanhua, T., Olsen, A., Hoppema, M., Jutterström, S., Schirnick, C., Van Heuven, S., Velo, A., Lin, X., Kozyr, A., Álvarez, M., Bakker, D.C.E., Brown, P., Falck,

- E., Jeansson, E., Lo Monaco, C., Olafsson, J., Pérez, F.F., Pierrot, D., Ríos, A.F., Sabine, C.L., Schuster, U., Steinfeldt, R., Stendardo, I., Anderson, L.G., Bates, N.R., Bellerby, R.G.J., Blindheim, J., Bullister, J.L., Gruber, N., Ishii, M., Johannessen, T., Jones, E.P., Köhler, J., Körtzinger, A., Metzl, N., Murata, A., Musielewicz, S., Omar, A.M., Olsson, K.A., de la Paz, M., Pfeil, B., Rey, F., Rhein, M., Skjelvan, I., Tilbrook, B., Wanninkhof, R., Mintrop, L., Wallace, D.W.R., Key, R.M., 2009. CARINA Data Synthesis Project. ORNLCDIAC-157 NDP-091 Carbon Dioxide Inf. Anal. Cent. Oak Ridge Natl. Lab. US Dep. Energy Oak Ridge Tenn. 37831-6335
- Tanhua, T., Steinfeldt, R., Key, R.M., Brown, P., Gruber, N., Wanninkhof, R., Pérez, F., Körtzinger, A., Velo, A., Schuster, U., Van Heuven, S., Bullister, J.L., Stendardo, I., Hoppema, M., Olsen, A., Kozyr, A., Pierrot, D., Schirnack, C., Wallace, D., 2010a. Atlantic Ocean CARINA data: overview and salinity adjustments. *Earth Syst. Sci. Data* 2, 17–34, DOI: 10.5194/essd-2-17-2010
- Tanhua, T., van Heuven, S., Key, R.M., Velo, A., Olsen, A., Schirnack, C., 2010b. Quality control procedures and methods of the CARINA database. *Earth Syst Sci Data* 2, 35–49, DOI: 10.5194/essd-2-35-2010, 2010
- Tans, P., Keeling, R., 2014. Trends in Atmospheric Carbon Dioxide [WWW Document]. Trends Atmospheric Carbon Dioxide. URL <http://www.esrl.noaa.gov/gmd/ccgg/trends/>
- Touratier, F., Azouzi, L., Goyet, C., 2007. CFC-11, D14C and 3H tracers as a means to assess anthropogenic CO₂ concentrations in the ocean. *Tellus B* 59, 318–325, DOI: 10.1111/j.1600-0889.2006.00247.x
- U.S. Department of Commerce, N.O. and A.A., 2006. 2-minute Gridded Global Relief Data (ETOPO2v2)
- Van Heuven, S., Pierrot, D., Lewis, E., Wallace, D., 2009. MATLAB Program developed for CO₂ system calculations. ORNL/CDIAC-105b, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee
- Vázquez-Rodríguez, M., Padín, X.A., Ríos, A.F., Bellerby, R.G.J., Pérez, F.F., 2009a. An upgraded carbon-based method to estimate the anthropogenic fraction of dissolved CO₂ in the Atlantic Ocean. *Biogeosciences Discuss* 6, 4527–4571, DOI: 10.1016/j.jmarsys.2011.11.023
- Vázquez-Rodríguez, M., Pérez, F.F., Velo, A., Ríos, A.F., Mercier, H., 2012. Observed acidification trends in North Atlantic water masses. *Biogeosciences* 9, 5217–5230, DOI: 10.5194/bg-9-5217-2012.
- Vázquez-Rodríguez, M., Touratier, F., Lo Monaco, C., Waugh, D.W., Padin, X.A., Bellerby, R.G.J., Goyet, C., Metzl, N., Ríos, A.F., Pérez, F.F., 2009b. Anthropogenic carbon distributions in the Atlantic Ocean: data-based estimates from the Arctic to the Antarctic. *Biogeosciences* 6, 439–451, DOI: 10.5194/bg-6-439-2009
- Velo, A., Pérez, F.F., Brown, P., Tanhua, T., Schuster, U., Key, R.M., 2009. CARINA alkalinity data in the Atlantic Ocean. *Earth Syst. Sci. Data* 1, 45–61

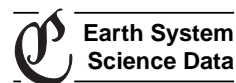
- Velo, A., Pérez, F.F., Lin, X., Key, R.M., Tanhua, T., de la Paz, M., Olsen, A., van Heuven, S., Jutterström, S., Ríos, A.F., 2010a. CARINA data synthesis project: pH data scale unification and cruise adjustments. *Earth Syst. Sci. Data* 2, 133–155, DOI: 10.5194/essd-2-133-2010
- Velo, A., Pérez, F.F., Tanhua, T., Gilcoto, M., Ríos, A.F., Key, R.M., 2013. Total alkalinity estimation using MLR and neural network techniques. *J. Mar. Syst.* 111–112, 11–18, DOI: 10.1016/j.jmarsys.2012.09.002
- Velo, A., Vázquez-Rodríguez, M., Padín, X.A., Gilcoto, M., Ríos, A.F., Pérez, F.F., 2010b. A multiparametric method of interpolation using WOA05 applied to anthropogenic CO₂ in the Atlantic. *Sci. Mar.* 74, 21–32
- Wallace, D.W., 1995. Monitoring global ocean carbon inventories. Ocean Observing System Development Panel
- Wanninkhof, R., Doney, S.C., Bullister, J.L., Levine, N.M., Warner, M., Gruber, N., 2010. Detecting anthropogenic CO₂ changes in the interior Atlantic Ocean between 1989 and 2005. *J. Geophys. Res. Oceans* 115, C11028, DOI: 10.1029/2010JC006251
- Waugh, D.W., Haine, T.W.N., Hall, T.M., 2004. Transport times and anthropogenic carbon in the subpolar North Atlantic Ocean. *Deep Sea Res. Part Oceanogr. Res. Pap.* 51, 1475–1491, DOI: 10.1016/j.dsr.2004.06.011
- Wolf-Gladrow, D.A., Zeebe, R.E., Klaas, C., Körtzinger, A., Dickson, A.G., 2007. Total alkalinity: The explicit conservative expression and its application to biogeochemical processes. *Mar. Chem.* 106, 287–300, DOI: 10.1016/j.marchem.2007.01.006
- Wootton, J., Pfister, C., Forester, J., 2008. Dynamic patterns and ecological impacts of declining ocean pH in a high-resolution multi-year dataset. *Proc. Natl. Acad. Sci.* 105, 18848
- Zeebe, R., Wolf-Gladrow, D., 2001. CO₂ in seawater: equilibrium, kinetics, isotopes. Elsevier Science

Apéndices

Capítulo I

CARINA alkalinity data in the Atlantic Ocean

Earth Syst. Sci. Data, 1, 45–61, 2009
 www.earth-syst-sci-data.net/1/45/2009/
 © Author(s) 2009. This work is distributed under
 the Creative Commons Attribution 3.0 License.



CARINA alkalinity data in the Atlantic Ocean

A. Velo¹, F. F. Perez¹, P. Brown², T. Tanhua³, U. Schuster², and R. M. Key⁴

¹Instituto de Investigaciones Marinas, CSIC, Eduardo Cabello, 6, 36208 Vigo, Spain

²School of Environmental Sciences, University of East Anglia, Norwich, NR4 7TJ, UK

³Leibniz-Institut für Meereswissenschaften, Marine Biogeochemie, Kiel, Germany

⁴Princeton University, Program in Atmospheric and Oceanic Science, Forrestal Campus/Sayre Hall, Princeton, NJ 08544, USA

Received: 25 June 2009 – Published in Earth Syst. Sci. Data Discuss.: 3 August 2009

Revised: 13 November 2009 – Accepted: 16 November 2009 – Published: 27 November 2009

Abstract. Data on carbon and carbon-relevant hydrographic and hydrochemical parameters from previously non-publicly available cruise data sets in the Arctic, Atlantic and Southern Ocean have been retrieved and merged to a new database: CARINA (CARbon IN the Atlantic).

These data have gone through rigorous quality control (QC) procedures to assure the highest possible quality and consistency. The data for most of the measured parameters in the CARINA data base were objectively examined in order to quantify systematic differences in the reported values, i.e. secondary quality control. Systematic biases found in the data have been corrected in the data products, i.e. three merged data files with measured, calculated and interpolated data for each of the three CARINA regions; Arctic, Atlantic and Southern Ocean. Out of a total of 188 cruise entries in the CARINA database, 98 were conducted in the Atlantic Ocean and of these, 75 cruises report alkalinity values.

Here we present details of the secondary QC on alkalinity for the Atlantic Ocean part of CARINA. Procedures of quality control, including crossover analysis between cruises and inversion analysis of all crossover data are briefly described. Adjustments were applied to the alkalinity values for 16 of the cruises in the Atlantic Ocean region. With these adjustments the CARINA database is consistent both internally as well as with GLODAP data, an oceanographic data set based on the World Hydrographic Program in the 1990s. Based on our analysis we estimate the internal accuracy of the CARINA-ATL alkalinity data to be $3.3 \mu\text{mol kg}^{-1}$. The CARINA data are now suitable for accurate assessments of, for example, oceanic carbon inventories and uptake rates and for model validation.

Data coverage and parameter measured

Repository-Reference: doi:10.3334/CDIAC/otg.CARINA.ATL.V1.0

Available at:

http://cdiac.ornl.gov/ftp/oceans/CARINA/CARINA_

Database/CARINA.ATL.V1.0/

Coverage: 60° S–75° N; 80° W–34° E

Location Name: Atlantic Ocean

Date/Time Start: 1977-10-7

Date/Time End: 2006-02-02



Correspondence to: A. Velo
 (avelo@iim.csic.es)

Published by Copernicus Publications.

Data Product Parameter Name	Data Product Flag name	Exchange File Parameter Name	Exchange File Flag Name	Units
station		STANBR		
day		DATE		
month		DATE		
year		DATE		
latitude		LATITUDE		decimal degrees
longitude		LONGITUDE		decimal degrees
cruisenro				
depth				meters
temperature		CTDTMP		°C
salinity	sf	SALNTY	SALNTY_FLAG.W	
pressure		CTDPRS		decibars
alk	alkf	ALKALI	ALKALI_FLAG.W	micromole kg ⁻¹

For a complete list of parameters for the CARINA data base, see Key et al. (2009). Note the different names for the parameters in the Exchange files (the individual cruise files) and the merged data product.

1 Introduction

CARINA is a database of carbon and carbon relevant data from hydrographic cruises in the Arctic, Atlantic and Southern Oceans. The project was formed as an essentially informal, unfunded project in Kiel, Germany, in 1999, with the main goal to create a database of carbon relevant variables in the ocean to be used for accurate assessments of oceanic carbon inventories and uptake rates. Not only the collection of data, but also the quality control of the data has been a main focus of the project, with both primary and secondary quality control (QC) of the data having been performed. The CARINA database consists of essentially two parts:

The first part are the individual cruise files where all the measured data, and their quality flags, are stored. These files are in WHP (WOCE Hydrographic Program) exchange format where the first lines consist of the condensed metadata. There are essentially no calculated neither interpolated values in the individual cruise files, with the exceptions of pressure calculated from depth and some bottle salinities that were taken from ctdsal. No adjustments have been applied to any of these values, with the exception that all pH measurements were converted to the seawater pH scale at 25°C.

The second part of CARINA are three merged quality controlled and adjusted data files; one each for the Atlantic Ocean, Arctic Mediterranean Seas and Southern Ocean regions. These files contain all the CARINA data and include: 1) interpolated values for nutrients, oxygen and salinity if those data are missing and if interpolation could be made according to criteria described in Key et al. (2009) (this special issue); calculated carbon parameters (e.g. if total dissolved inorganic carbon (TCO₂) and Total Alkalinity (*A_T*) were measured, pH can be calculated). Calculated and interpolated values have the quality flag "0". All the values in the merged data file have been adjusted according to the values in Table 1 and described in Sect. 5. In many cases there are more reported parameters in the individual cruise files that

has been included in the secondary QC, such as ¹⁴C, ¹³C and SF₆.

This report describes the consistency analysis of alkalinity measurements of the Atlantic Ocean part of the CARINA database (CARINA-ATL). A more comprehensive description of the complete CARINA data base can be found in Key et al. (2009) (this special issue), for an overview of the North Atlantic CARINA data, see Tanhua et al. (2009a) (this special issue). Alkalinity reports for other areas included in CARINA can be found in the following reports: Arctic Ocean by Jutterström et al. (2009), Nordic Seas by Olsen et al. (2009), Atlantic Sector of Southern Ocean by Hoppema et al. (2009), Indian Sector of Southern Ocean by Lo Monaco et al. (2009) and Pacific Sector of Southern Ocean by Sabine et al. (2009).

The Total Alkalinity in sea water is defined as "...the number of moles of hydrogen ion equivalent to the excess of proton acceptors (bases formed from weak acids with a dissociation constant $K \leq 10^{-4.5}$, at 25°C and zero ionic strength) over proton donors (acids with $K > 10^{-4.5}$) in one kilogram of sample." (Dickson, 1981). *A_T* is one of the four basic related parameters of the carbon dioxide "CO₂" system in seawater, with the others being total dissolved inorganic carbon (*C_T*), the fugacity of dissolved CO₂ (*f*CO₂), and pH. If at least two of these are known, then the remaining parameters can be calculated and the entire CO₂ system determined using thermodynamic constants for a given temperature, salinity and pressure.

High quality seawater carbon data are critical for detecting small changes in the CO₂ system. Both *A_T* and *C_T* are used in a number of methods for calculating the anthropogenic CO₂ signal, most specifically those that employ back-calculation techniques. *A_T* is a key in both determining changes in *C_T* produced by CaCO₃ dissolution (Feely et al., 2002) and establishing concentrations of *C_T* in surface waters at equilibrium with the atmosphere, whether at historical, present, or future CO₂ levels. Furthermore, many measurements of the CO₂ system in seawater have been performed using the

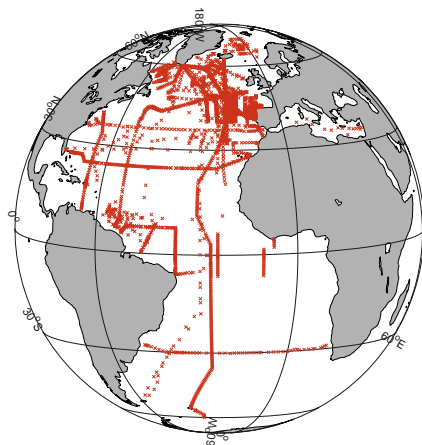


Figure 1. Map of stations with alkalinity data in the CARINA-ATL dataset.

A_T -pH pair for determining C_T . For these reasons, this is a key parameter for the CARINA objectives.

2 Data

Alkalinity data included in the CARINA-ATL dataset originates from a multitude of international research groups using a number of different analysis methods. Whilst most of the A_T data was determined using closed cell potentiometric titrations (Dickson et al., 2007), many measurements were also made by potentiometric titration at end-point in open cell (Mintrop et al., 2000).

The CARINA-ATL dataset has a total of 52 043 alkalinity samples, coming from 4080 stations and 63 cruises. Of these, 46 961 samples came from measurements, and were flagged with a “2”. The remaining 5082 samples came from carbon calculations with the final adjusted data, and so were flagged with a “0”. Figure 1 shows the location of the stations with alkalinity data in CARINA-ATL.

For consistency with historical data, a further 12 WOCE/GLODAP reference cruises were included, giving a total of 75 cruises and 65 531 samples. Primary quality control – consisting of outlier and scatter identification – led to 2 cruises and 5738 samples being discarded. For the crossover analyses, only data deeper than 1500 m were used, leading to a number of cruises being omitted from the analyses. In total, 12828 CARINA-ATL samples and 2878 WOCE/GLODAP samples are compared here.

Of the 53 cruises included in the crossover analysis, alkalinity data from 31 had been generated using certified reference materials (CRMs) to test or calibrate the titration system. The remaining 22 did not use CRM at all, instead using an alternative solution for standardization (Dickson et al., 2007). The most typical reported analytical error was around 1%, $\sim 2.5 \mu\text{mol kg}^{-1}$. However, similar crossover exercises (Key et al., 2004; Sabine et al., 2005; Wanninkhof et al., 2003) performed on older data estimated an overall accuracy of $\pm 5 \mu\text{mol kg}^{-1}$.

The overall objective of this work was to assess the quality of the Atlantic CARINA alkalinity data in order to generate a mutually consistent database, by checking for apparent offsets for each cruise considering all the information reported by different sources (Tanhua et al., 2009b) For A_T , the minimum adjustment to be applied to the final dataset following the identification of an additive offset was established as $\pm 6 \mu\text{mol kg}^{-1}$, as was the case for the previous crossover comparison study for the North Atlantic (Wanninkhof et al., 2003).

3 Methods

The methods and techniques applied here are described in detail in Tanhua et al. (2009b). In summary, the procedure essentially entails the comparison of data from separate cruises whose tracks cross or at least come close to each other, so-called crossover analyses. Before that comparison, a primary quality control (1st QC) consisting on outlier and scatter identification has been performed (Key et al., 2009). Only data that were flagged “good” during this primary QC procedure were considered in this process. With the application of various software packages (Tanhua et al., 2009b) generating statistical and objective information about the offsets between pairs of cruises, as well as the graphics needed to visually verify the computer generated offsets. For each crossover analysis, A_T data from samples deeper than 1500 m were compared on sigma-4 density surfaces generating an offset and a standard deviation of this difference, as well as totaling the number of contributing stations and samples. For alkalinity, additive offsets (and adjustments) were determined.

Slightly different procedures have been performed in the Nordic Seas (Olsen, 2009) due to the lower variability in deep waters in that area.

In this work, the semi-automated crossover procedure was run for all possible pairs of Atlantic cruises. Next, the crossover results were then visually inspected in order to ensure quality and to check the analysis had run correctly. Only “good” quality crossovers were selected, and those results were used for subsequent cruise adjustment calculations. Good crossovers had enough sample data to yield a reasonably uniform additive offset over the entire zone of analysis, leading to parallel cruise profiles. Standard deviations for

48

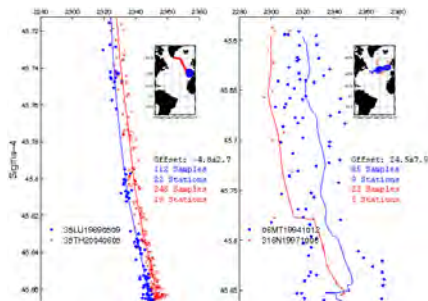


Figure 2. Crossover between the cruises 35LU19890509 and 35TH20040605 (a) and the cruises 06MT19941012 and 316N19971006 (b).

individual cruise data and difference profiles were also used to provide more information on crossover quality.

In Fig. 2, two examples of A_T crossovers are shown; the contrast between a good crossover (on the left) and a bad one (on the right) can be easily appreciated, as well as the lack of data and dispersion. The quality of the offset is assessed when the standard deviation of the offset is compared (± 2.7 versus $\pm 7.9 \mu\text{mol kg}^{-1}$).

After this first iteration, automated procedures described by Tanhua et al. (2009b) weighted the crossover quality by statistical parameters for later adjustment calculations for all cruises. This process was applied for all available cruise crossovers in the Atlantic. In total, 337 individual crossovers were obtained for A_T .

The final offsets and statistics were used as the input for an inverse least squares procedure (Tanhua et al., 2009b). Inversion results generated a set of suggested *corrections* for all cruises included in the analysis that minimized the differences. Figure 3 shows the offsets for all A_T crossovers in the Atlantic region before (pink dots), and after (blue dots) the adjustments were applied to each cruise. The convergence to values inside the bounds can be easily appreciated for values after inversions. The standard error for the original crossovers is $10.5 \mu\text{mol kg}^{-1}$ whereas the standard error for the adjusted ones is $4.7 \mu\text{mol kg}^{-1}$.

In order to ensure the highest quality results from the inversion and to help get a more accurate and consistent solution to the system, a small subset of cruises were a priori defined as “core”. These were chosen according to their geographical extent (i.e. covering a large distance) and expected high data quality (i.e. WOCE/CLIVAR quality), and were agreed upon by the CARINA Atlantic group. Offsets identified towards “core” cruises received a higher weighting in the inversion minimization process (Tanhua et al., 2009b). Once the full result of A_T offsets for each cruise had been

A. Velo et al.: CARINA alkalinity data in the Atlantic Ocean

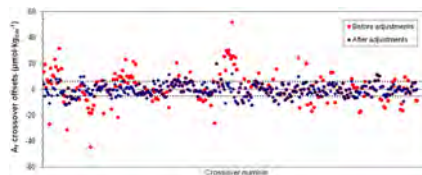


Figure 3. Crossover alkalinity offsets obtained with original database and after adjustments were applied.

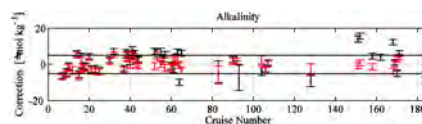


Figure 4. Mean and standard deviation of the offset for each cruise before (black) and after (red) the adjustments were applied.

generated, only the suggested corrections that exceeded a predefined limit of $\pm 6 \mu\text{mol kg}^{-1}$ were applied. These subsets of high values were used as starting point to establish the final adjustment values to be proposed. The decision on final values was made manually and by consensus among the CARINA collaborators. For this decision, subjective factors such as the location, use of CRMs, technique, date, quantity and quality of crossovers, or some particularly relevant crossovers were used. Only those offsets that were strongly supported by the analysis were finally adopted and subsequently applied to the measured results. The corrections that were actually applied to the data product are, in following, referred to as an *adjustment*.

After the need for an adjustment and its magnitude were established for each cruise, the values were applied to the original database, and the full process of crossovers and inversions repeated. The result of this second iteration was a very useful way of validating the proposed offsets. Figure 4 shows the corrections values for each cruise obtained after the first inversion procedures (values in black), and the corrections obtained after the second iteration (values in red). The results from the second inversion clearly show that the remaining offsets are lower, and most of them fit within the chosen minimum error boundary

3.1 Overall accuracy

The offsets for the crossovers applied to the data product were used to estimate the overall accuracy of the alkalinity data (Fig. 5). The weighted mean (WM) was calculated for alkalinity by using the absolute value of the offset (D) of the

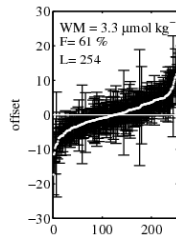


Figure 5. Sorted offsets calculated for the crossovers in the CARINA-ATL data after adjustments have been applied. *WM*: the weighted mean of the offsets (see text); *F*: the percentage of offsets indistinguishable from 1 within their uncertainty; *L*: the number of crossovers.

L crossovers with the uncertainty (σ):

$$WM = \frac{\sum_{i=1}^L D(i)/(\sigma(i)^2)}{\sum_{i=1}^L 1/(\sigma(i)^2)} \quad (1)$$

Based on this analysis we have estimated the accuracy of the CARINA-ATL alkalinity data to $3.3 \mu\text{mol kg}^{-1}$.

4 Results

Results are summarized in Table 1. This table shows the cruises in the CARINA-ATL dataset with A_T data, and the summary of adjustments applied. The following data is presented:

- Cruise ID: CARINA assigned identification number for the cruise.
- Cruise Expocode: String identifying the cruise. This is composed by a country code (two numbers), vessel code (two characters or numbers) and the departure date in year, month, day format (YYYYMMDD).
- Region: Location of the cruise. All cruises belong to Atlantic Ocean (NA), but some overlap with the Arctic Mediterranean Seas (AMS) or Southern Ocean (SO) areas.
- Core: Indicates whether or not the cruise is a core cruise for the crossover analysis.
- CRM: Indicates whether or not CRMs were used during instrumental analysis.
- WLSQ adj: Result of the inversion process through the Weighted Least Squares method (Tanhua et al., 2009b).

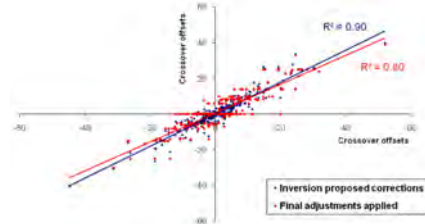


Figure 6. Original (x-axis) versus post-adjust (y-axis) offsets for the all crossover in the Atlantic Ocean after applying the full solution (blue) or the final applied solution given in the last column of Table 1 (pink).

- WDSQ adj: Result of the inversion process through the Weighted Damped Least Squares method (Tanhua et al., 2009b).
- Adjustment: Adjustments applied for the cruises in the merged data product. All adjustments are fully supported by the CARINA group and no adjustments smaller than $\pm 6 \mu\text{mol kg}^{-1}$ are applied.

Figure 6 is a comparison between the original corrections and the final adjustments applied. Blue dots represent offsets obtained by applying full solution (all corrections); pink dots represent offsets obtained with the final adjustments applied to the data product (Table 1). The relation for the applied adjustments (pink dots) has only a slightly lower correlation coefficient than for the full solution.

In the following paragraph a set of figures and comments are presented for each cruise summarizing all crossover offsets with their standard deviation. Each figure shows the following information:

- Green dots: “Offsets”. These values are the offsets taken directly from each crossover. The standard deviation is shown as error bars on these dots.
- Yellow line indicates the additive correction calculated by inversions for the cruise. Note that the correction and offsets are of opposite sign.
- Black stars indicate the correction calculated by inversions for the other cruises that intersect this cruise.
- Blue squares: “Predicted offset” shows the calculated offset that would be obtained by applying all inversion corrections to the cruises.
- Red dots: These are the residuals between the “Offsets” (Green dots) and “Predicted Offsets” (Blue squares)
- c suffix in the upper X-axis labels stands for Core Cruises.

Table 1. CARINA-ATL dataset with alkalinity data and adjustments applied.

Cruise ID	Expocode	Region	Core	CRM	WLSQ adj \pm STD	WDLSQ adj \pm 95%CI	Adjustment $\mu\text{mol kg}^{-1}$
7	06BE20001128	NA					
8	06GA19960613	NA					
9	06GA20000506	NA					
10	06MT19920316	NA					
12	06MT19920509	NA		x	-27.9 ± 1.3	-25.1 ± 1.6	-25
14	06MT19920701	NA+AMS					
15	06MT19940219	NA					
16	06MT19941012	NA					
17	06MT19941115	NA					
18	06MT19960613	NA					
20	06MT19960910	NA					
21	06MT19970107	NA					
23	06MT19970515	NA					
25	06MT19970707	NA					
26	06MT19970815	NA					
28	06MT19990610	NA					
30	06MT19990711	NA					
32	06MT19990813	NA					
44	06MT20010507	NA	x	x	7.4 ± 0.8	6.2 ± 1.2	6
51	06MT20010620	NA					
52	06MT20010717	NA			6.7 ± 1.0	7.1 ± 1.0	7
53	06MT20011018	NA					
54	06MT20020607	NA					
55	06MT20021013	NA			6.7 ± 0.9	6.5 ± 1.3	6
56	06MT20030626	NA					
57	06MT20030723	NA			4.8 ± 0.9	4.3 ± 1.7	6
60	06MT20030831	NA					
61	06MT20040311	NA		x	-3.9 ± 1.8	-5.1 ± 0.8	-6
62	18HU19920527	NA					
63	18HU19930405	NA					
64	18HU19930617	NA			-10.2 ± 0.9	-9.6 ± 1.1	-10
65	18HU19931105	NA			5.5 ± 0.8	8.1 ± 0.9	8
66	18HU19940524	NA					
68	18HU19941012	NA					
69	18HU19950419	NA					
84	18HU19970509	NA					
85	29CS19771007	NA					
86	29CS19930510	NA					
87	29GD19821110	NA					
89	29GD19831201	NA					
90	29GD19840218	NA					
92	29GD19840711	NA					
93	29GD19860904	NA					
94	29HE19980730	NA					
95	29HE20010305	NA+SO					
106	29HE20020304	NA+SO		x	-4.6 ± 2.6	-6.8 ± 1.6	-6
107	29HE20030408	NA					
108	31AN19890420	NA					
109	316N19971005	NA					
113	316N20010627	NA					
125	316N20030922	NA					
153	316N20031023	NA			13.7 ± 0.9	13.8 ± 0.7	14
154	32EV19910328	NA			15.2 ± 0.7	14.4 ± 1.4	14
158	32OC19950529	NA					
160	33LK19960415	NA			4.1 ± 0.7	5.2 ± 0.7	5
164	33RO19980123	NA					
165	33RO20030604	NA					
168	33RO20050111	NA+SO					
170	33SW20010102	NA			11.9 ± 0.4	14.9 ± 1.6	15
171	33SW20030418	NA					
172	34AR19970805	NA+AMS	x	x	-5.3 ± 0.6	-8.5 ± 0.6	-9
173	35A320010203	NA					
188	35A320010322	NA			-9.3 ± 1.0	-8.1 ± 2.0	-8
	323019940104	NA+SO	x	x	-10.9 ± 2.8	-8.9 ± 0.8	-8

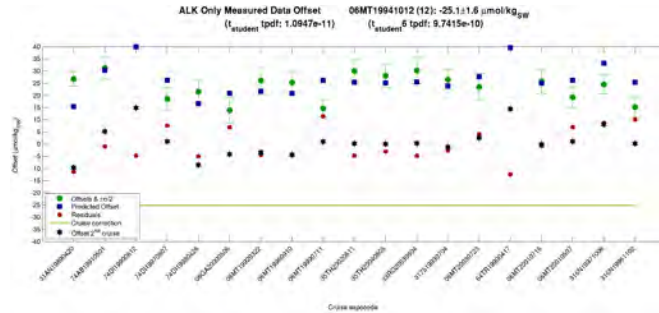


Figure 7. Cruise crossover information plot for 06MT19941012.

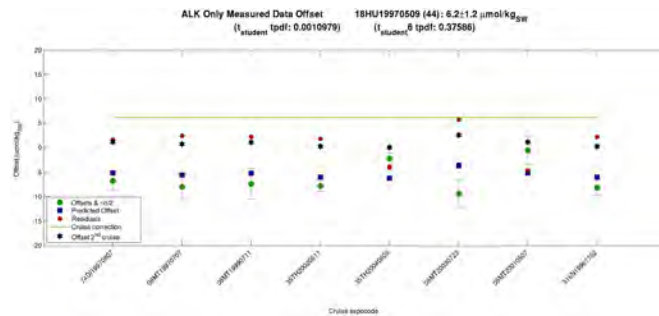


Figure 8. Cruise crossover information plot for 18HU19970509.

5 Cruises

In this section, an assessment and description of the adjustments applied to cruises for CARINA-ATL database is made. CARINA identifiers for the cruises are the numbers indicated between the parentheses. Exact data locations can be found at the CARINA website: http://cdiac.ornl.gov/oceans/CARINA/Carina_inv.html.

5.1 Cruise 06MT19941012 (12) (Fig. 7)

This is the so called MT30/2 cruise, on board Meteor and along WOCE leg A02 section, in the North Atlantic. It has 53 stations taken with a 24 place rosette system. A closed cell potentiometric titration method was used on measurements, and CRM Batch #22 was used as reference. The report indicates an estimated precision of $\pm 3 \mu\text{mol kg}^{-1}$ and an accuracy $\pm 6 \mu\text{mol kg}^{-1}$. This cruise has 19 crossovers. The inversions suggest a correction of $-25.1 \mu\text{mol kg}^{-1}$. Most of

the residuals fit within $\pm 10 \mu\text{mol kg}^{-1}$, with half part of them inside $\pm 5 \mu\text{mol kg}^{-1}$ after corrections applied. No good fit exists with the core-cruises however. The crossovers with 74DI19970807 and 06GA20000506 suggest a minor offset, but the other four core cruises (two of them GLODAP), suggest a correction of around $-28 \mu\text{mol kg}^{-1}$. Based on this evidence, an adjustment of $-25 \mu\text{mol kg}^{-1}$ was applied to the alkalinity data.

5.2 Cruise 18HU19970509 (44) (Fig. 8)

This is the AR07Wh, a CCHDO cruise on board R/V Hudson, in the Labrador Sea. It has 13 stations taken with a 24 place Rosette system with ten-liter bottles. CRM analyses were used for standardization and precision was $2.7 \mu\text{mol kg}^{-1}$. There are eight crossovers giving a fitted correction of $6.2 \mu\text{mol kg}^{-1}$. The suggested adjustment for this cruise is therefore $6 \mu\text{mol kg}^{-1}$. Most of the residuals fit very close to 0 and keep inside $\pm 5 \mu\text{mol kg}^{-1}$ after offsets applied.

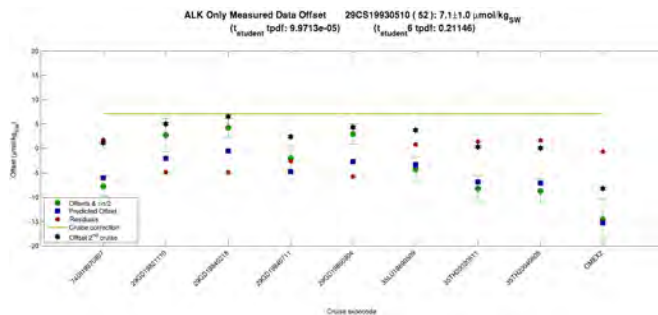


Figure 9. Cruise crossover information plot for 29CS19930510.

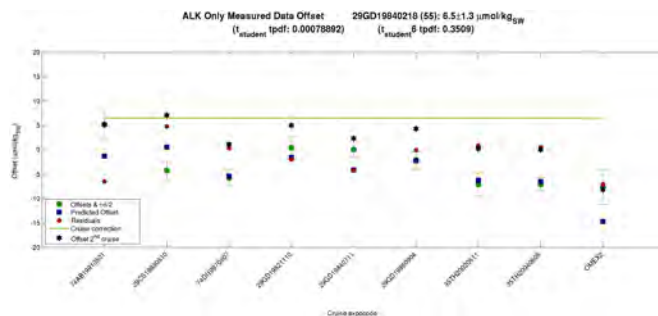


Figure 10. Cruise crossover information plot for 29GD19840218.

There are very good fits with four core cruises. Based on this evidence, an adjustment of $6 \mu\text{mol kg}^{-1}$ was applied to the alkalinity data.

5.3 Cruise 29CS19930510 (52) (Fig. 9)

This is the so called MORENA-I cruise conducted on board R/V Cornide de Saavedra, along WOCE line AR16e in the Atlantic area, close to NW Spain. It has 92 stations and 24 sampling levels using a rosette system. CRMs were not used for these measurements. There are nine crossovers. The fitted correction of inversions is $7.1 \mu\text{mol kg}^{-1}$. The suggested adjustment is therefore $7 \mu\text{mol kg}^{-1}$. Crossovers with three core cruises show very low residuals after offset applied. Based on this evidence, an adjustment of $7 \mu\text{mol kg}^{-1}$ was applied to the alkalinity data.

5.4 Cruise 29GD19840218 (55) (Fig. 10)

This is the so called GALICIA-VII cruise, on board R/V Garcia del Cid in the Atlantic area close to NW Spain. It has 33 stations taken on a hydrocast with 1.71 Niskin bottles. A precision of 0.1% and accuracy of $1.4 \mu\text{mol kg}^{-1}$ has been reported. No CRMs were used during the analysis. There are nine crossovers giving a fitted correction of $6.5 \mu\text{mol kg}^{-1}$. The suggested adjustment is $+6 \mu\text{mol kg}^{-1}$. Almost all residuals fit very close to 0 and stay within $\pm 5 \mu\text{mol kg}^{-1}$ after offsets applied. There are very good fits with three core cruises. Based on this evidence, an adjustment of $6 \mu\text{mol kg}^{-1}$ was applied to the alkalinity data.

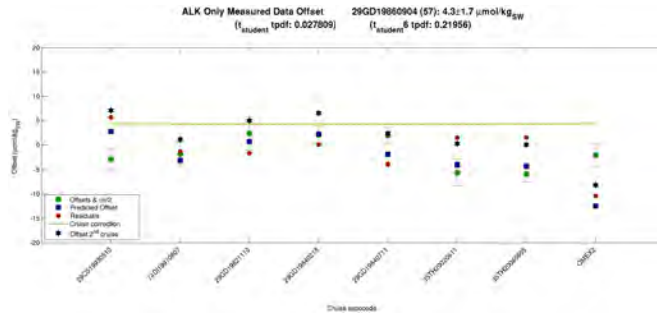


Figure 11. Cruise crossover information plot for 29GD19860904.

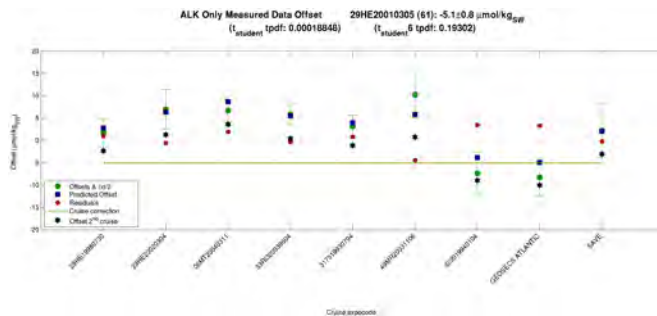


Figure 12. Cruise crossover information plot for 29HE20010305.

5.5 Cruise 29GD19860904 (57) (Fig. 11)

This is the so called GALICIA-IX cruise, on board R/V Garcia del Cid in the Atlantic area close to NW Spain. It has 50 stations taken on hydrocasts with 1.7L Niskin bottles. A precision of 0.1‰ and accuracy of $1.4 \mu\text{mol kg}^{-1}$ has been reported. No CRMs were used.

There are nine crossovers giving a fitted correction of $+4.3 \mu\text{mol kg}^{-1}$. The suggested adjustment is $+6 \mu\text{mol kg}^{-1}$ as t-student statistical checks show that the proposed offset is indistinguishable from $6 \mu\text{mol kg}^{-1}$. Almost all residuals fit very close to zero and keep inside $\pm 5 \mu\text{mol kg}^{-1}$ after offsets applied. There are very good fits with three core cruises. Based on this evidence, an adjustment of $6 \mu\text{mol kg}^{-1}$ was applied to the alkalinity data.

5.6 Cruise 29HE20010305 (61) (Fig. 12)

This is the so called FICARAM II cruise, conducted on R/V Hesperides along WOCE section A17 in the Western South Atlantic area. It has 29 full depth stations taken with a 24 place rosette system. CRM batches 41 and 51 were used. Uncertainty is $1.4 \mu\text{mol kg}^{-1}$ according to cruise report. There are 9 crossovers, two of them in the Southern Ocean giving a fitted correction of $-5.1 \mu\text{mol kg}^{-1}$. Analysis of cruise documentation seems to suggest a correction of $-5 \mu\text{mol kg}^{-1}$. The proposed adjustment is $-6 \mu\text{mol kg}^{-1}$, as it is within the offset error and is better supported than no adjustment. All residuals are very close to 0 and fit inside $\pm 5 \mu\text{mol kg}^{-1}$. There are very good fits with five core cruises. Based on this evidence, an adjustment of $-6 \mu\text{mol kg}^{-1}$ was applied to the alkalinity data.

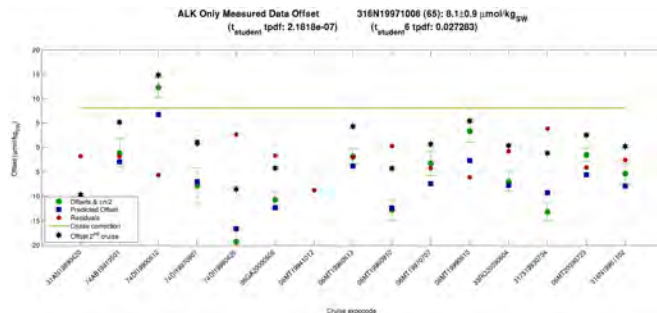


Figure 13. Cruise crossover information plot for 316N19971005.

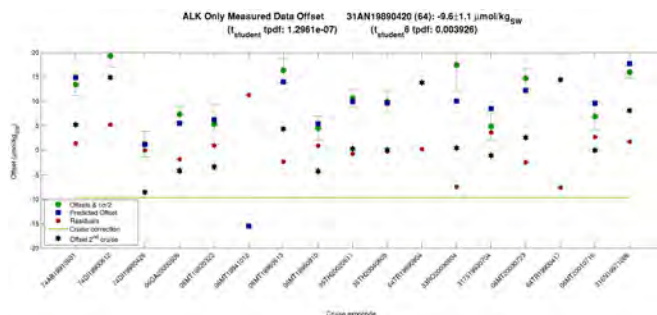


Figure 14. Cruise crossover information plot for 31AN19890420.

5.7 Cruise 316N19971005 (65) (Fig. 13)

This is the so called KN154/2 cruise, on board R/V Knorr along the WOCEAR24b section in the NW Atlantic area. It has 162 stations and 24 sampling levels using a rosette system. No reference to the use of CRM was reported. There are 15 crossovers giving a fitted correction of $+8.1 \mu\text{mol kg}^{-1}$. The suggested adjustment is $+8 \mu\text{mol kg}^{-1}$. Excepting one, all inversion residuals fit inside $\pm 6 \mu\text{mol kg}^{-1}$. The cruise has a good fit with three core-cruises. Based on this evidence, an adjustment of $8 \mu\text{mol kg}^{-1}$ was applied to the alkalinity data.

5.8 Cruise 31AN19890420 (64) (Fig. 14)

This is the so called 31A119 cruise, on board R/V Atlantis II and conducted along a section at about 20°W from 47°N to 60°N . This cruise has 51 stations taken with a rosette system. According to the cruise report, the values seem to be about 5 to $10 \mu\text{mol kg}^{-1}$ too high in relation to OACES93 and A16N2003. CRMs were not used. The fitted correction is $-9.6 \mu\text{mol kg}^{-1}$. The inversions suggest an adjustment of $-10 \mu\text{mol kg}^{-1}$. Crossovers with core cruises support the proposed offset very well. In addition, GLODAP 317519930704 supports this adjustment. Based on this evidence, an adjustment of $-10 \mu\text{mol kg}^{-1}$ was applied to the alkalinity data.

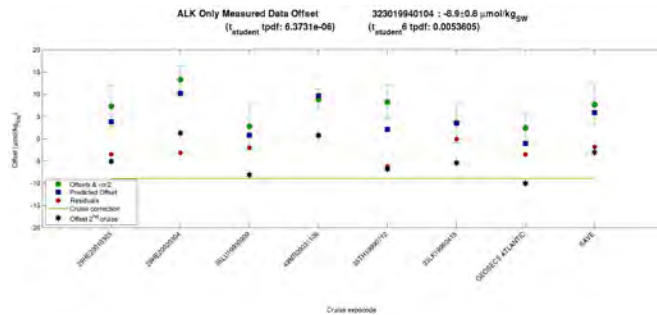


Figure 15. Cruise crossover information plot for 323019940104.

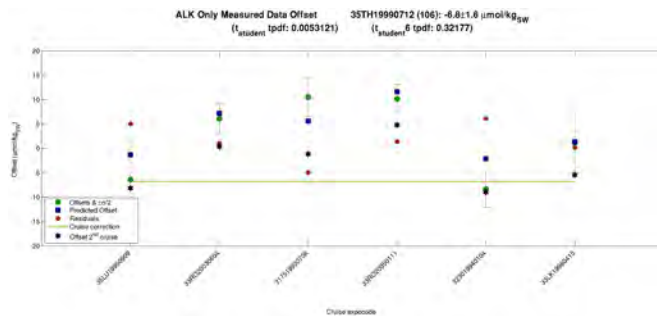


Figure 16. Cruise crossover information plot for 35TH19990712.

5.9 Cruise 323019940104 (Fig. 15)

This is the CITHER2 cruise, on board R/V Maurice Ewing and along the WOCE A17 section in the Western South Atlantic area from 10° S to 55° S. It has 235 stations taken with a 32 place rosette system. In the cruise report, an offset for alkalinity of $-8 \mu\text{mol kg}^{-1}$ was stated. There are eight crossovers that give a fitted correction of $-8.9 \mu\text{mol kg}^{-1}$. The suggested adjustment is $-8 \mu\text{mol kg}^{-1}$ because it agrees with both the fitted result and the cruise report. Most of the inversion residuals fit inside $\pm 5 \mu\text{mol kg}^{-1}$. There are very good fits with four GLODAP cruises. Based on this evidence, an adjustment of $-8 \mu\text{mol kg}^{-1}$ was applied to the alkalinity data.

5.10 Cruise 35TH19990712 (106) (Fig. 16)

This is the so called EQUALANT99 cruise, on board R/V Thalassa in the Equatorial Atlantic area. It has 102 stations taken with a 24 place rosette system. Precision is reported as $1.7 \mu\text{mol kg}^{-1}$. A comparison with WOCE A15 data implies offset values as high as $12.7 \mu\text{mol kg}^{-1}$. There are six crossovers, giving a fitted correction of $-6.8 \mu\text{mol kg}^{-1}$. The suggested adjustment is $-6 \mu\text{mol kg}^{-1}$. All inversion residuals fit inside $\pm 6 \mu\text{mol kg}^{-1}$ after corrections are applied. There are very good fits with two GLODAP cruises. Based on this evidence, an adjustment of $-6 \mu\text{mol kg}^{-1}$ was applied to the alkalinity data.

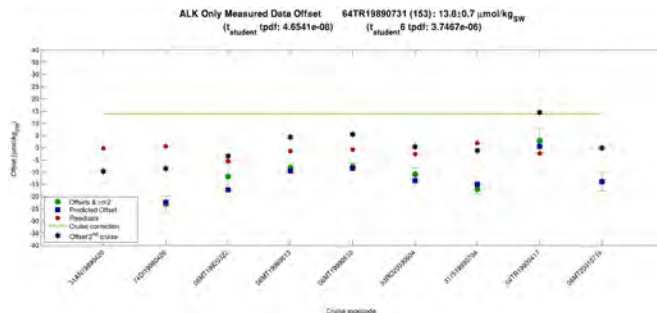


Figure 17. Cruise crossover information plot for 64TR19890731.

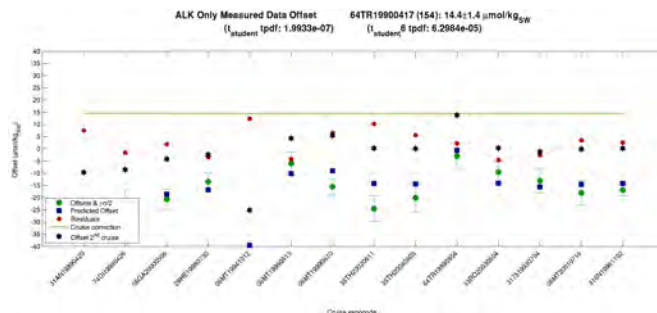


Figure 18. Cruise crossover information plot for 64TR19900417.

5.11 Cruise 64TR19890731 (153) (Fig. 17)

This cruise is also called 64TY8908, on board R/V Tyro along a meridional A16N section at 20° W. The cruise has 73 stations and 12 sampling levels using a rosette system. Nine stations have been flagged 3. There are nine crossovers giving a fitted correction of $13.8 \mu\text{mol kg}^{-1}$. The suggested adjustment is $14 \mu\text{mol kg}^{-1}$. All inversion residuals are close to zero and fit inside the $\pm 5 \mu\text{mol kg}^{-1}$ boundary. There are very good fits with two GLODAP cruises. Based on this evidence, an adjustment of $14 \mu\text{mol kg}^{-1}$ was applied to the alkalinity data.

5.12 Cruise 64TR19900417 (154) (Fig. 18)

This cruise is also called 64TY9001, on board R/V Tyro along a meridional A16N section at 20° W from 30° N to 60° N. This cruise has 23 stations with multiple casts on most of them. The method of Bradshaw et al. (1981), and calculations with constants of Goyet and Poisson (1989) were used. Values are about $20 \mu\text{mol kg}^{-1}$ low relative to CLIVAR A16N-2003 and have twice the scatter. There are 14 crossovers. The fitted correction is $+14.4 \mu\text{mol kg}^{-1}$. The suggested adjustment is $+14 \mu\text{mol kg}^{-1}$. After fitted corrections are applied, 10 inversion residuals fit inside $\pm 5 \mu\text{mol kg}^{-1}$. There are good fits with two GLODAP and one core cruise (06GA20000506). Based on this evidence, an adjustment of $14 \mu\text{mol kg}^{-1}$ was applied to the alkalinity data.

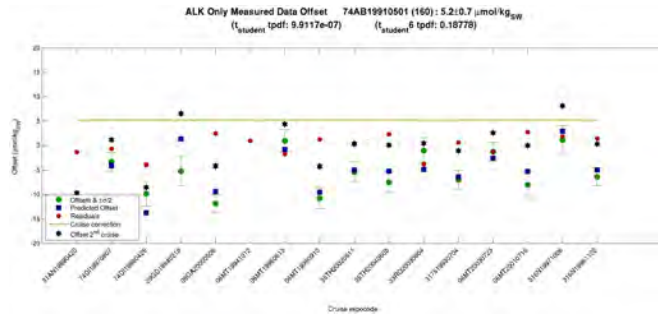


Figure 19. Cruise crossover information plot for 74AB19910501.

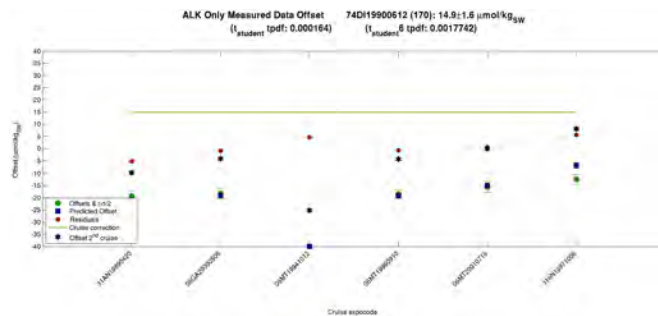


Figure 20. Cruise crossover information plot for 74DI19900612.

5.13 Cruise 74AB19910501 (160) (Fig. 19)

This is the so called Vivaldi expedition, on board R/V Charles Darwin in the NE Atlantic area primarily West of Iberia. It has 614 stations, from which only 34 are deep stations. The samples were taken with a 24 place rosette system. According to the cruise report, the values are low in relation to GLODAP by about $15 \mu\text{mol kg}^{-1}$. Good precision is noted. Laboratory-made borax standards were used but no CRMs. There are 16 crossovers giving a fitted adjustment of $5.2 \mu\text{mol kg}^{-1}$. The suggested adjustment is $5 \mu\text{mol kg}^{-1}$ as a student t-test shows that the proposed correction is indistinguishable from 6. Crossovers with six core cruises also support the proposed offset. Furthermore, GLODAP 317519930704 supports this adjustment, as very low residuals are present after the offset has been applied. Based on this evidence, an adjustment of $5 \mu\text{mol kg}^{-1}$ was applied to the alkalinity data.

5.14 Cruise 74DI19900612 (170) (Fig. 20)

This cruise is also called 74DI192, and was conducted on R/V Discovery in the North Atlantic area from 46°N to 49°N and 17°W to 15°W . This cruise has 20 stations and 12 sampling levels using a rosette system. Comparing with one GLODAP station, values are about $18 \mu\text{mol kg}^{-1}$ lower. There are six crossovers giving a fitted correction of $14.9 \mu\text{mol kg}^{-1}$. The inversions suggest an adjustment of $15 \mu\text{mol kg}^{-1}$. Inversion residuals are inside a $\pm 5 \mu\text{mol kg}^{-1}$ boundary after corrections are applied. Based on this evidence, an adjustment of $15 \mu\text{mol kg}^{-1}$ was applied to the alkalinity data.

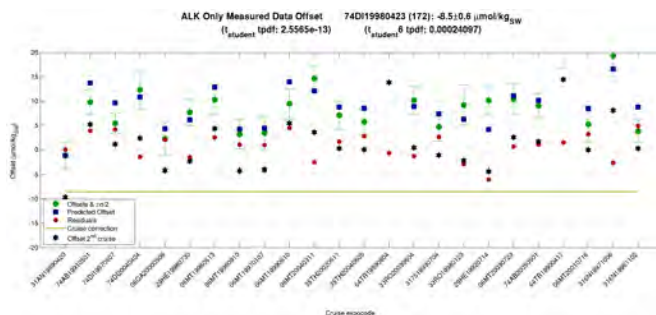


Figure 21. Cruise crossover information plot for 74DI19980423.

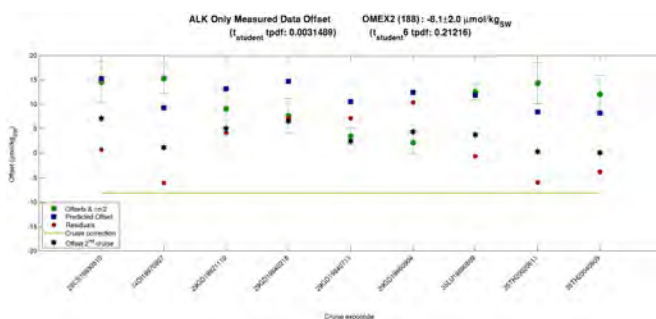


Figure 22. Cruise crossover information plot for OMEX2.

5.15 Cruise 74DI19980423 (172) (Fig. 21)

This is the 74DI233 cruise, also called Chaos, conducted on R/V Discovery, performing a meridional section along 20° W from 20° N to 60° N. It has 44 full depth stations taken with a 24 place rosette system. Data is generally good, with a few data high relative to neighbours and GLODAP. CRMs were used. There are 24 crossovers giving a fitted correction of $-8.5 \mu\text{mol kg}^{-1}$. The suggested adjustment is $-9 \mu\text{mol kg}^{-1}$. Most of the residuals fit very close to zero and keep inside $\pm 5 \mu\text{mol kg}^{-1}$ after corrections are applied. There are very good fits with core and GLODAP cruises. Based on this evidence, an adjustment of $-9 \mu\text{mol kg}^{-1}$ was applied to the alkalinity data.

5.16 Cruise OMEX2 (188) (Fig. 22)

This is also called OMEGX2 Project Time Series. Rather than a single cruise, it is a compilation of 12 cruises performed on board R/Vs Belgica, Charles Darwin and Meteor on the NW Iberian margin. There are nine crossovers giving a fitted correction of $-8.1 \mu\text{mol kg}^{-1}$. The suggested adjustment is $-8 \mu\text{mol kg}^{-1}$. With one exception, all inversion residuals fit inside $\pm 6 \mu\text{mol kg}^{-1}$. There are three crossovers with core cruises that appear to suggest an even larger offset. Based on this evidence, an adjustment of $-8 \mu\text{mol kg}^{-1}$ was applied to the alkalinity data.

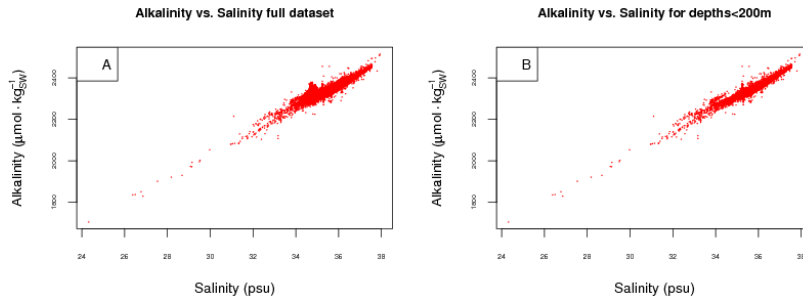


Figure 23. Alkalinity data in CARINA final database versus salinity for: (A) data deeper than 200 m (B) depths <200 m.

6 Data quality evaluation

As derived from its definition, the salt content of seawater is a determining factor on the A_T . For this reason, salinity is a good parameter to explain some of the A_T variability. In Fig. 23a, alkalinity is shown against salinity for the full CARINA-ATL dataset. As expected, the parameters correlate well, giving a R^2 regression coefficient of 0.77. As near surface formation of water masses is the main process that affects A_T variability, Fig. 23b, represents the variation of A_T versus salinity for the subset of surface waters (depth <200 m). The correlation for this data set is higher than in Fig. 23a, giving a R^2 of 0.94 due to the fact that most of the variability in both salt and alkalinity occurs in the upper few hundred meters. This enforces the idea of the interdependence of both parameters.

To make an overall evaluation of the dataset quality, additional variables that can affect the natural variability of the alkalinity were introduced in the regression. Thus, a Multi-Linear Regression (MLR) was performed with potential temperature (θ), salinity, latitude, apparent oxygen utilisation (AOU), nitrate, phosphate and silicate in order to remove as much natural variability as possible. All of these parameters were included knowing that interdependence exists between many of them. However, the goal here is not to generate a statistical model for alkalinity, but rather a way of getting residuals with natural variability removed to a large extent (about 90%), so that they can be used to test the data quality. This analysis method facilitates a better assessment of scatter and biases in the alkalinity dataset, and even to test the applied adjustments. The MLR procedure can transmit the measurement error of the explanatory variables to the residuals despite being quite lower than the alkalinity error.

In order to improve the quality of the evaluation, the MLR analysis was applied in four density layers, with density at 1000 db (σ_1) being used to divide the ocean. The upper thermocline was set by $\sigma_1 < 32.25 \text{ kg m}^{-3}$; intermediate waters

(depths from about 1000 to 2000 m) were defined by second layer ($32.25 \leq \sigma_1 < 32.39 \text{ kg m}^{-3}$); water depths between ~2000 to 3000 m were defined by $32.39 \leq \sigma_1 < 32.53 \text{ kg m}^{-3}$ – corresponding to North Atlantic Deep Waters (NADW); and finally, the fourth layer refers to bottom waters, where the presence of Antarctic Bottom Waters (AABW) dominates. This last layer is set by $\sigma_1 > 32.53 \text{ kg m}^{-3}$. The surface layer with depths <200 m was removed from this evaluation.

Using this analysis, alkalinity residuals were calculated by the following Eq. (2):

$$A_{T_MLR} = \sum_{i=1}^8 a_i \cdot X_i \quad (2)$$

$$A_{T_residuals} = A_{T_measured} - A_{T_MLR}$$

where X_i stand for θ , salinity, latitude, AOU, nitrate, phosphate, silicate, and a constant term. This procedure was done with the CARINA-ATL corrected database, and also for the database without any alkalinity adjustments applied.

The alkalinity residuals for each density layer are shown in Fig. 24. This shows a box plot of alkalinity residuals cruise data for each cruise. The box/whiskers represent the typical five number summary (minimum, 25th percentile, median, 75th percentile and maximum), the width of individual boxes corresponds to the number of samples for that cruise, and the numbers on the x-axis are the CARINA cruise IDs as referenced in Table 1.

The best fit ($R^2=0.95$) is obtained for the shallower waters, with a mean standard deviation of $6.2 \mu\text{mol kg}^{-1}$. The other layers have a slightly lower mean standard deviation (5.4 , $5.9 \mu\text{mol kg}^{-1}$ respectively), except for the bottom layer with $6.5 \mu\text{mol kg}^{-1}$. The median of mean standard deviations for all cruises is $4.1 \mu\text{mol kg}^{-1}$ units for each of three deepest density layers. The lower panel in the figure stands for the joined alkalinity residuals of the four density layers.

As can be seen, alkalinity residuals are lower when using the corrected database, in comparison with the uncorrected

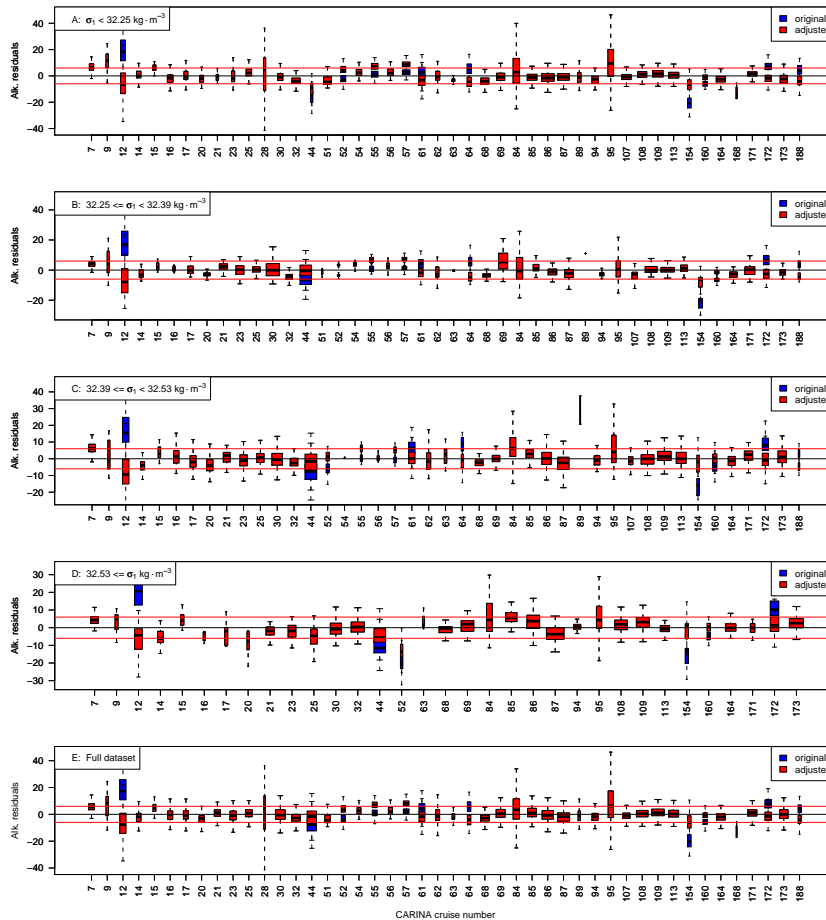


Figure 24. Alkalinity residuals obtained from CARINA-ATL dataset by applying an MLR for alkalinity data against Theta, Salinity, Latitude, AOU, Nitrate, Phosphate and Silicate. (A to D) are CARINA-ATL subsets for the indicated σ_t interval, and (E) is the join for the full dataset. Blue values are residuals with the original unadjusted alkalinity values, and Red values are the final adjusted alkalinity values. Red lines are the $\pm 6 \mu\text{mol kg}^{-1}$ of alkalinity used as lower limit for adjustments in the crossover exercise.

original ones. Most but not all of the cruises have the alkalinity residuals median inside of the $\pm 6 \mu\text{mol kg}^{-1}$ boundary, showing that in addition to the alkalinity measurement errors, there are two other sources that increase the variability of the alkalinity residuals: firstly, the MLR is not able to ex-

plain all of the naturally observed variability of alkalinity; and secondly, the predictor parameters have their own inherent measurement errors.

Acknowledgements. This work has been performed and funded as part of the EU project CarboOcean (Project 511176). Additional support from the International Ocean Carbon Coordination Project IOCCP (Maria Hood) and the Hanse Institute for Advanced Study (HWK Delmenhorst, Germany) are gratefully acknowledged. Additional support provided as follows: for A. Velo, F. F. Perez, PGIDIT05OXIC40203PM Xunta de Galicia and CTM200627116E/MAR MEC; P. J. Brown, Award NER/S/A/2004/12255; and R. M. Key, NOAA grant NA08OAR4320752 and NA08OAR4310820. We want also to thank the hard work and dedication of analysts and investigators that collect the data at the sea; without them, this work would not be possible. We gratefully acknowledge also those who have contributed their data to the CARINA project.

Edited by: M. Hoppema

References

- Bradshaw, A., Brewer, P., Shafer, D., and Williams, R.: Measurements of total carbon dioxide and alkalinity by potentiometric titration in the geosecs program, *Earth Planet. Sci. Lett.*, 55, 99–115, 1981.
- Dickson, A.: An exact definition of total alkalinity and a procedure for the estimation of alkalinity and total inorganic carbon from titration data, *Deep-Sea Res.*, 28, 609–623, 1981.
- Dickson, A. G., Sabine, C. L., and Christian, J. R.: Guide to best practices for ocean CO₂ measurements, *PICES Special Publication*, 3, 191 pp., 2007.
- Feely, R. A., Sabine, C. L., Lee, K., Millero, F. J., Lamb, M. F., Greeley, D., Bullister, J. L., Key, R. M., Peng, T. H., Kozyr, A., Ono, T., and Wong, C. S.: In situ calcium carbonate dissolution in the Pacific ocean, *Global Biogeochem. Cy.*, 16, 1144, doi:10.1029/2002GB001866, 2002.
- Goyet, C. and Poisson, A.: New determination of carbonic acid dissociation constants in seawater as a function of temperature and salinity, *Deep-Sea Res.*, 36, 1635–1654, 1989.
- Hoppema, M., Velo, A., van Heuven, S., Tanhua, T., Key, R. M., Lin, X., Bakker, D. C. E., Perez, F. F., Ríos, A. F., Lo Monaco, C., Sabine, C. L., Álvarez, M., and Bellerby, R. G. J.: Consistency of cruise data of the CARINA database in the Atlantic sector of the Southern Ocean, *Earth Syst. Sci. Data Discuss.*, 2, 331–365, 2009, <http://www.earth-syst-sci-data-discuss.net/2/331/2009/>.
- Jutterström, S., Anderson, L. G., Bates, N. R., Bellerby, R., Johannessen, T., Jones, E. P., Key, R. M., Lin, X., Olsen, A., and Omar, A. M.: Arctic Ocean data in CARINA, *Earth Syst. Sci. Data Discuss.*, 2, 281–308, 2009, <http://www.earth-syst-sci-data-discuss.net/2/281/2009/>.
- Key, R. M., Kozyr, A., Sabine, C. L., Lee, K., Wanninkhof, R., Bullister, J. L., Feely, R. A., Millero, F. J., Mordy, C., and Peng, T. H.: A global ocean carbon climatology: Results from global data analysis project (glodap), *Global Biogeochem. Cy.*, 18, GB4031, doi:10.1029/2004GB002247, 2004.
- Key, R. M., Tanhua, T., Olsen, A., Hoppema, M., Jutterström, S., Schirnack, C., van Heuven, S., Kozyr, A., Lin, X., Velo, A., Wallace, D. W. R., and Mintrop, L.: The CARINA data synthesis project: Introduction and overview, *Earth Syst. Sci. Data Discuss.*, in press, 2009.
- Lo Monaco, C., Álvarez, M., Key, R. M., Lin, X., Tanhua, T., Tilbrook, B., Bakker, D. C. E., van Heuven, S., Hoppema, M., Metzl, N., Ríos, A. F., Sabine, C. L., and Velo, A.: Assessing the internal consistency of the CARINA database in the Indian sector of the Southern Ocean, *Earth Syst. Sci. Data Discuss.*, 2, 367–419, 2009, <http://www.earth-syst-sci-data-discuss.net/2/367/2009/>.
- Mintrop, L., Pérez, F., González-Dávila, M., Santana-Casiano, J., and Körtzinger, A.: Alkalinity determination by potentiometry: Inter calibration using three different methods, *Ciencias Marinas*, 26, 23–37, 2000.
- Olsen, A.: Nordic Seas total alkalinity data in CARINA, *Earth Syst. Sci. Data Discuss.*, 2, 309–330, 2009, <http://www.earth-syst-sci-data-discuss.net/2/309/2009/>.
- Sabine, C., Key, R., Kozyr, A., Feely, R., Wanninkhof, R., Millero, F., Peng, T., Bullister, J., and Lee, K.: Global ocean data analysis project (glodap): Results and data, *ndp-083*, 110 pp., Carbon Dioxide Inf. Anal. Cent., Oak Ridge Natl. Lab., Oak Ridge, Tenn., 2005.
- Sabine, C. L., Hoppema, M., Key, R. M., Tilbrook, B., van Heuven, S., Lo Monaco, C., Metzl, N., Ishii, M., Murata, A., and Musielewicz, S.: Assessing the internal consistency of the CARINA data base in the Pacific sector of the Southern Ocean, *Earth Syst. Sci. Data Discuss.*, 2, 555–578, 2009, <http://www.earth-syst-sci-data-discuss.net/2/555/2009/>.
- Tanhua, T., Steinfeldt, R., Key, R. M., Brown, P., Gruber, N., Wanninkhof, R., Perez, F., Körtzinger, A., Velo, A., Schuster, U., van Heuven, S., Bullister, J. L., Stendardo, I., Hoppema, M., Olsen, A., Kozyr, A., Pierrot, D., Schirnack, C., and Wallace, D. W. R.: Atlantic Ocean CARINA data: overview and salinity adjustments, *Earth Syst. Sci. Data Discuss.*, 2, 241–280, 2009a, <http://www.earth-syst-sci-data-discuss.net/2/241/2009/>.
- Tanhua, T., van Heuven, S., Key, R. M., Velo, A., Olsen, A., and Schirnack, C.: Quality control procedures and methods of the CARINA database, *Earth Syst. Sci. Data Discuss.*, 2, 205–240, 2009b, <http://www.earth-syst-sci-data-discuss.net/2/205/2009/>.
- Wanninkhof, R., Tsung-Hung, P., Huss, B., Sabine, C. L., and Lee, K.: Comparison of inorganic carbon system parameters measured in the Atlantic ocean from 1990 to 1998 and recommended adjustments, *Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, ORNL/CDIAC-140*, 2003.

Capítulo I

**CARINA data synthesis project: pH data scale
unification and cruise adjustments**

Earth Syst. Sci. Data, 2, 133–155, 2010
 www.earth-syst-sci-data.net/2/133/2010/
 doi:10.5194/essd-2-133-2010
 © Author(s) 2010. CC Attribution 3.0 License.

Open Access
 Earth System
 Science
 Data

CARINA data synthesis project: pH data scale unification and cruise adjustments

A. Velo¹, F. F. Pérez¹, X. Lin², R. M. Key², T. Tanhua³, M. de la Paz¹, A. Olsen^{6,5}, S. van Heuven⁴, S. Jutterström⁵, and A. F. Ríos¹

¹Instituto de Investigaciones Marinas, CSIC, Eduardo Cabello, 6, 36208 Vigo, Spain

²Atmospheric and Oceanic Sciences Program, Princeton University, Princeton, NJ, USA

³Leibniz-Institute für Meereswissenschaften, Marine Biogeochemie, Kiel, Germany

⁴Netherlands Institute for Sea Research, Texel, The Netherlands

⁵Institutionen för kemi, Göteborgs universitet, Göteborg, Sweden

⁶Bjerknes Centre for Climate Research, UNIFOB AS, Allégaten 55, 5007 Bergen, Norway

Received: 19 August 2009 – Published in Earth Syst. Sci. Data Discuss.: 13 October 2009

Revised: 19 April 2010 – Accepted: 22 April 2010 – Published: 11 May 2010

Abstract. Data on carbon and carbon-relevant hydrographic and hydrochemical parameters from 188 previously non-publicly available cruise data sets in the Arctic Mediterranean Seas (AMS), Atlantic Ocean and Southern Ocean have been retrieved and merged to a new database: CARINA (CARbon IN the Atlantic Ocean).

These data have gone through rigorous quality control (QC) procedures to assure the highest possible quality and consistency. The data for most of the measured parameters in the CARINA database were objectively examined in order to quantify systematic differences in the reported values. Systematic biases found in the data have been corrected in the data products, three merged data files with measured, calculated and interpolated data for each of the three CARINA regions; AMS, Atlantic Ocean and Southern Ocean. Out of a total of 188 cruise entries in the CARINA database, 59 reported pH measured values. All reported pH data have been unified to the Sea-Water Scale (SWS) at 25 °C.

Here we present details of the secondary QC of pH in the CARINA database and the scale unification to SWS at 25 °C. The pH scale has been converted for 36 cruises. Procedures of quality control, including crossover analysis between cruises and inversion analysis are described. Adjustments were applied to the pH values for 21 of the cruises in the CARINA dataset. With these adjustments the CARINA database is consistent both internally as well as with the GLODAP data, an oceanographic data set based on the World Hydrographic Program in the 1990s. Based on our analysis we estimate the internal consistency of the CARINA pH data to be 0.005 pH units. The CARINA data are now suitable for accurate assessments of, for example, oceanic carbon inventories and uptake rates, for ocean acidification assessment and for model validation.

Data coverage and parameter measured

Repository-Reference:

doi:10.3334/CDIAC/otg.CARINA.AMS.V1.2

doi:10.3334/CDIAC/otg.CARINA.ATL.V1.0

doi:10.3334/CDIAC/otg.CARINA.SO.V1.0

Available at:

[http://cdiac.ornl.gov/ftp/oceans/CARINA/CARINA_](http://cdiac.ornl.gov/ftp/oceans/CARINA/CARINA_Database/CARINA.AMS.V1.2/)

[Database/CARINA.AMS.V1.2/](http://cdiac.ornl.gov/ftp/oceans/CARINA/CARINA_Database/CARINA.AMS.V1.2/)

[http://cdiac.ornl.gov/ftp/oceans/CARINA/CARINA_](http://cdiac.ornl.gov/ftp/oceans/CARINA/CARINA_Database/CARINA.ATL.V1.0/)

[Database/CARINA.ATL.V1.0/](http://cdiac.ornl.gov/ftp/oceans/CARINA/CARINA_Database/CARINA.ATL.V1.0/)

[http://cdiac.ornl.gov/ftp/oceans/CARINA/CARINA_](http://cdiac.ornl.gov/ftp/oceans/CARINA/CARINA_Database/CARINA.SO.V1.0/)

[Database/CARINA.SO.V1.0/](http://cdiac.ornl.gov/ftp/oceans/CARINA/CARINA_Database/CARINA.SO.V1.0/)

Coverage: 78° S–90° N; 180° W–180° E

Location Name: Arctic Mediterranean Seas, Atlantic Ocean and Southern Ocean

Date/Time Start: 1977-10-7

Date/Time End: 2006-02-10



Correspondence to: A. Velo
 (avelo@iim.csic.es)

Published by Copernicus Publications.

Data Product Parameter Name	Data Product Flag Name	Exchange File Parameter Name	Exchange File Flag Name	Units
station		STANBR		
day		DATE		
month		DATE		
year		DATE		
latitude		LATITUDE		decimal degree
longitude		LONGITUDE		decimal degree
cruisenr				
depth				meter
temperature		CTDTMP		°C
salinity	sf	SALNTY	SALNTY_FLAG_W	
pressure		CTDPRS		decibar
phws25	phws25f	PH_SWS	PH_SWS_FLAG_W	

For an introduction to this work and a complete list of parameters in the CARINA data base, see Key et al. (2010) and Tanhua et al. (2009). Note the different names for the parameters in the Exchange files (the individual cruise files) and the merged data product.

1 Introduction

Carbon-related data from both historical and recent hydrographic cruises in the Arctic Mediterranean Seas (AMS, includes Arctic Ocean and Nordic Seas), Atlantic and Southern Oceans have been brought together to form the CARINA database. The major aim of this project was to produce an internally consistent dataset of carbon-related parameters that can be used to assess and quantify carbon uptake and storage in these regions. Focus was placed not only on the collection of relevant data but also ensuring quality. The CARINA working group has performed both primary and secondary quality control (QC). This report is a summary of the pH data in the CARINA data set and describes the data consistency analysis (secondary QC) and scale conversions undertaken. For an introduction to and overview of the work done in the CARINA project see Key et al. (2010), Tanhua et al. (2010) as well as the other more specialized papers of this special issue.

1.1 Description of parameter (pH)

pH is one of the four parameters that define the carbonate system in sea water. The term pH describes the acidity of a liquid and it is defined as: $\text{pH} = -\log_{10} [\text{H}^+]$. The pH of seawater has become a valuable oceanographic parameter, particularly since problems with its measurement and interpretation have been resolved through the development of rational pH scales (Dickson, 1993), photometric measurement methods (Clayton and Byrne, 1993) and reliable pH buffer standards of seawater (Dickson, 1993; Millero et al., 1993).

Oceanic pH reflects the thermodynamics state of the acid-base system in seawater, especially of the geochemically important carbonate system. The equilibrium between carbonate species and CO_2 , the carbonate buffering, helps to mitigate to a large extent the changes that can be induced to seawater pH by several causes (Millero, 2007; Raven et al., 2005; Wootton et al., 2008). Uptake of CO_2 from the atmo-

sphere can incur changes in pH, but the carbonate buffering acts to stabilize these changes by consumption of carbonate ions present in the seawater, and ultimately by taking carbonate ions through dissolution of CaCO_3 sediments, shifting the equilibrium.

1.2 Distribution

The primary factors governing the spatial and temporal distribution of ocean pH are temperature, because of the dependence of the dissociation constant on this parameter, and partial pressure of CO_2 ($p\text{CO}_2$), total alkalinity (A_T) and total carbon (C_T), due to carbonate equilibrium. Surface ocean pH is mainly affected by temperature, biology, uptake/release of CO_2 from/to atmosphere, and uptake/release of CO_2 from/to deeper waters by upwelling or sinking. There are changes and cycles that contribute to fluctuations in the surface ocean pH, on timescales of days to years and including seasons (Raven et al., 2005; Wootton et al., 2008). It is estimated that the global surface ocean has already acidified by roughly 0.1 pH units (from 8.2 to 8.1 in SWS scale at 25 °C) since pre-industrial times (Caldeira and Wickett, 2003; Olafsson et al., 2009; Orr et al., 2005; Raven et al., 2005) due to anthropogenic CO_2 emission, and its uptake by the ocean. In the deep oceans, the CO_2 concentration increases by decomposition of organic matter that sinks, and these additions of CO_2 cause its pH to decrease, but because CaCO_3 is abundant in sediments, the pH of the deep oceans is very stable, and changes are minimal even on timescales of 10 000 years (Raven et al., 2005).

1.3 Different pH scales

The first pH definition ($\text{pH} = -\log_{10} [\text{H}^+]$) from Sørensen (1909) presents some operational problems since free protons $[\text{H}^+]$ do not exist in any significant amount in aqueous solutions. Thus, the symbol “ H^+ ” represents hydrate complexes rather than the concentration of free hydrogen

Table 1. pH scales used on seawater measurements, definitions, and relationships (Dickson, 1993; Dickson et al., 2007; Millero, 2007).

Scale	pH definition
NBS	$\text{pH}_{\text{NBS}} = -\log a_{\text{H}^+}$
Free	$\text{pH}_{\text{F}} = -\log [\text{H}^+]_{\text{F}}$
Total	$\text{pH}_{\text{T}} = -\log [\text{H}^+]_{\text{T}} = -\log([\text{H}^+]_{\text{F}} \cdot (1 + [\text{SO}_4^{2-}]/K_{\text{S}})) \approx -\log([\text{H}^+]_{\text{F}} + [\text{HSO}_4^-])$
Seawater	$\text{pH}_{\text{SWS}} = -\log [\text{H}^+]_{\text{SWS}} = -\log([\text{H}^+]_{\text{F}} \cdot (1 + [\text{SO}_4^{2-}]/K_{\text{S}} + [\text{F}^-]/K_{\text{F}})) = -\log([\text{H}^+]_{\text{F}} + [\text{HSO}_4^-] + [\text{HF}])$

ions. The first operational definition was the NBS pH scale (Bates and Vih, 1973). NBS pH scale is defined by a series of standard buffer solutions with assigned pH values close to the best estimates of the proton activity (a_{H^+}), so that $\text{pH}_{\text{NBS}} = -\log a_{\text{H}^+}$. The reference state for pH_{NBS} scale is the indefinitely diluted solution, which is very useful in dilute natural waters such as rivers and lakes. However, this scale is not recommended for seawater because of its large ionic strength (Dickson, 1984; Millero et al., 1993).

In addition to the NBS pH scale, three other scales have been suggested for seawater, the free hydrogen ion scale (pH_{F}), the total hydrogen ion scale (pH_{T}) and the seawater scale (pH_{SWS}). The reason for the existence of four simultaneous pH scales is primarily historical. They reflect the gradual refinement of the experimentally determined pH in seawater. The definitions of the different scales are summarized in Table 1.

The free pH scale is conceptually the clearest being explicitly defined only by the H^+ concentration. The drawback is that the H^+ cannot be directly measured. This operational inconvenience is resolved with the total and seawater scales. The total pH scale (Hansson, 1973) accounts for the dissociation of HSO_4^- ion (including SO_4^{2-} in its calibration solutions), avoiding the definition of the HSO_4^- dissociation constant whose accurate value is difficult to obtain in seawater. The seawater scale (Dickson and Riley, 1979) includes besides bisulphate, the dissociation reaction for hydrogen fluoride. The differences between the total and the seawater scales thus arise from the fact whether the medium in which the scale is based includes fluoride or not. However, this difference is numerically small (about ~ 0.01 pH units at salinity 35) because the concentration of HSO_4^- is much larger than that of HF in the seawater. Contrarily, the pH reported on the free scale is about 0.11 and 0.12 pH units higher than on the total and the seawater scale (Zeebe and Wolf-Gladrow, 2001). These differences are much larger than the present precision achieved in the pH measurements, which is on the order of ± 0.0004 – 0.001 pH units (Clayton and Byrne, 1993).

Confusion may arise when the pH scale is not explicitly stated, and significant errors can be introduced in the calculation of the carbonic acid speciation because the first and second dissociation constant of H_2CO_3 are defined for a specific pH scale; thus, if the pH scale is ignored, serious errors in the pCO_2 calculation can occur that can reach $100 \mu\text{atm}$, specially whenever pH is a master variable of the carbonic system.

1.4 Methodology of pH measurements

Two analysis techniques are routinely used to get precise measurements of pH in seawater. These are potentiometric methods with electrodes, and spectrophotometric methods with an indicator.

The potentiometric method is based on the hydrogen ion sensitivity of an electrode (Dickson, 1993). It has fewer requirements on equipment, but is prone to problems due to electrode drift, susceptibility to electromagnetic interferences or problems with reference electrodes (Dickson, 1993). The accuracy of the measurement relies on the preparation of the calibration buffers. These, together with the precision of the temperature control, are common sources of biases/noise. The reproducibility of the potentiometric method is no better than ± 0.02 pH units (Dickson, 1993).

Spectrophotometric methods are based on the absorbance of a pH indicator, thus eliminating problems associated with buffer preparation and handling. Furthermore, errors due to poor temperature control can be partially reduced by using m-cresol (meta-cresol) as the indicator, since its pK value is centered in the typical range of the oceanic pH, and the temperature dependence follows that of pH along this range (Clayton and Byrne, 1993; Friis et al., 2004). The reproducibility with this method can reach to ± 0.0004 pH units (Clayton and Byrne, 1993).

2 Data

The pH data included in the CARINA dataset comes from a recompilation of cruise data from a multitude of international research groups until 2005. The data includes measurements obtained using both the potentiometric and spectrophotometric methods. Totalizing, CARINA dataset has pH data from 3761 stations on 59 cruises, resulting in 49 915 pH measurements. Figure 1 shows the location of the stations with pH data in CARINA (all ATL, SO and AMS datasets joined).

3 Conversions to pH SWS 25 °C

As a first step, all pH values were converted to the seawater scale at 25 °C (SWS 25). The conversion was done by using the CO2SYS (Lewis et al., 1998) routines converted to MATLAB code (van Heuven et al., 2009). This toolbox can be accessed at <http://cdiac.ornl.gov/oceans/co2rprt.html> and does all calculations needed to get the full solution of the

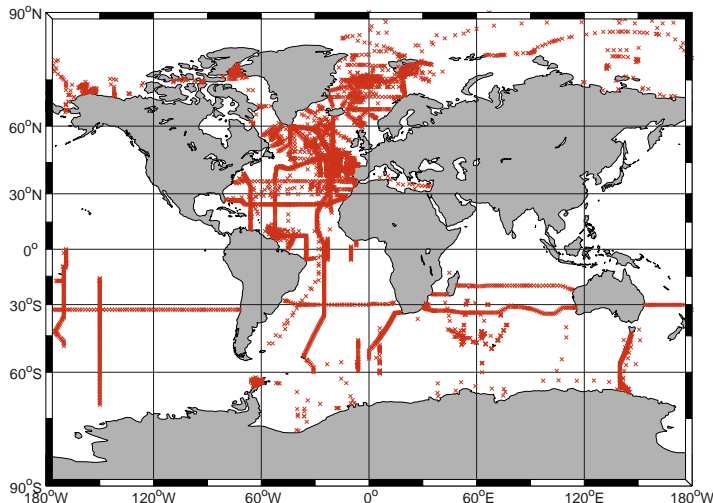


Figure 1. Map of stations with pH data in the full CARINA dataset (ATL, AMS and SO merged).

Table 2. Equations used by CO2SYS matlab software routines for conversion on pH scales. The routines use Total Scale internally for CO₂ calculations.

Equations used for pH scale conversion
$pH_T = pH_{SWS} - \log((1 + S_T/K_S)/(1 + S_T/K_S + F_T/K_F))$
$pH_T = pH_F - \log(1 + S_T/K_S)$
$pH_T = pH_{NBS} - (\log(1 + S_T/K_S) + \log(f_H))$

carbonate system in seawater. The code calculates the pH in all of the four scales used in seawater, and allows the user to choose which set of constants to use for the carbonate and sulfate systems.

Additional parameters required for these calculations are pressure, salinity, silicate, phosphate, and alkalinity. For the conversions, the following constants were used: Mehrbach refitted by Dickson and Millero (Dickson and Millero, 1987; Mehrbach et al., 1973) for carbonate, and Dickson (Dickson, 1990) for sulphate. The matlab routine converts all pH values to the total scale using the equations summarized in Table 2, which, in turn, are based on the definitions summarized in Table 1.

The next step taken by the CO2SYS routines is to calculate all four parameters of the CO₂ system (C_T, A_T, pH and

pCO₂) to the output temperature of 25 °C (still on the total scale). This is achieved by recalculating the constants, solving the system, and calculating the pH again.

As the last step, these values are used to calculate the pH on the other three scales. The output in pH_{SWS} is calculated from pH_T by reversing the first equation in Table 2:

$$pH_{SWS} = pH_T + \log \left(\frac{1 + \frac{S_T}{K_S}}{1 + \frac{S_T}{K_S} + \frac{F_T}{K_F}} \right) \quad (1)$$

Where: S_T is the Total Sulfate, calculated with the equations from Morris and Riley (1966), F_T is the Total Fluorine, calculated with the equations from Riley (1965), K_S is the bisulfate ion dissociation constant, from Dickson (1990), K_F is the hydrogen fluoride dissociation constant, from Dickson (1979), and f_H is the activity coefficient of hydrogen ion in seawater (Pérez and Fraga, 1987), calculated according to Takahashi et al. equations (1982).

Table 3 shows the cruises identified to be in a scale different than SWS at 25 °C. Original scale is noted in the second column and the reported temperature in the third.

4 Methods

The methods and techniques for the quality control are described in detail in the methods paper of this special issue (Tanhua et al., 2010). This secondary quality control (2nd

Table 3. Scale conversions. Note for cruise 165: Temperature specified for each sample, between 16.42 and 22.04 °C; Note for cruise 187: data quality low.

Cruise ID	Cruise Expocode	Scale	Temperature (°C)
3	06AQ19960712	T	15
20	06MT19990610	T	22
21	06MT19990711	T	22
23	06MT20010507	T	21
25	06MT20010717	T	21
30	06MT20030723	T	21
51	29CS19771007	NBS	15
52	29CS19930510	NBS	15
58	29HE19951203	NBS	15
59	29HE19960117	NBS	15
61	29HE20010305	T	25
62	29HE20020304	T	25
84	33LK19960415	T	in situ T/P
92	35A320010203	T	25
93	35A320010322	T	25
95	35LU19950909	T	in situ T/P
99	35MF19990104	T	in situ T
107	35TH20010823	T	25
108	35TH20020611	T	25
109	35TH20040604	T	25
110	49HH19941213	NBS	25
115	49ZS19921203	NBS	in situ T
119	58AA19950217	T	15
141	58JH19970414	T	15
142	58JH19980801	T	15
148	58LA19860719	T	25
160	74AB19910501	NBS	15
163	74AB20020301	T	25
165	74DI19890511	SWS	lab
172	74DI19980423	T	25
168	74DI19900425	NBS	25
170	74DI19900612	NBS	25
179	77DN20020420	T	15
182	90AV20041104	T	25
183	91AA19971204	T	15
187	OMEX1NA	NBS	in situ T

QC) starts after the unification of the pH data to SWS scale at 25 °C for all cruises. Next, an overview of the terminology is presented for improved readability.

- Crossover: Comparison of any parameter between a pair of cruises (A and B) located near or that cross each other. The comparison is done by station profiles.
- Offset: The numeric result of the crossover. Can be additive or multiplicative, and represents the quantity that cruise B data is biased (offset) from cruise A data. The crossover procedure returns also a statistical uncertainty for the offset.

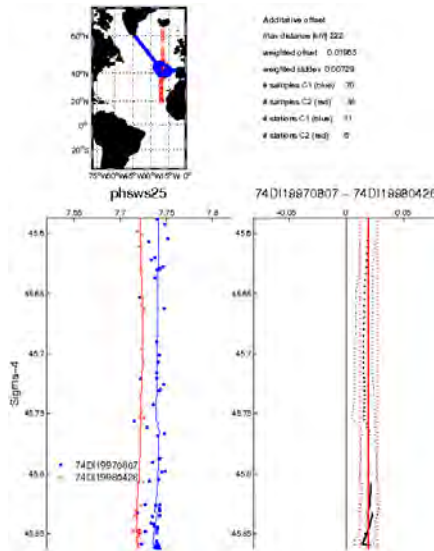


Figure 2. Crossover made with Running-Cluster Routines.

- Inversions: Weighted and Weighted Damped Least Squares (WLSQ and WDLQ) procedure. It uses the collection of offsets and their uncertainties as input, and calculates the optimal corrections for the individual cruise files needed in order to minimize the offsets.
- Corrections: The value of correction for each cruise directly returned by the inversions.
- Minimum adjustment: Not all corrections proposed by the inversions were applied. A minimum threshold was defined for each parameter. Below this limit, no adjustment was performed.
- Adjustments: Final adjustment value applied to the cruise. The correction given by inversions was taken as reference, but manual supervision and agreement was done by the CARINA group for each adjustment.

4.1 Crossovers (offsets)

The secondary QC procedure starts with comparisons of data from pairs of cruises that are either co-located or “near” each other. Only data that were flagged “good” during the primary QC procedure (Key et al., 2010) were considered.

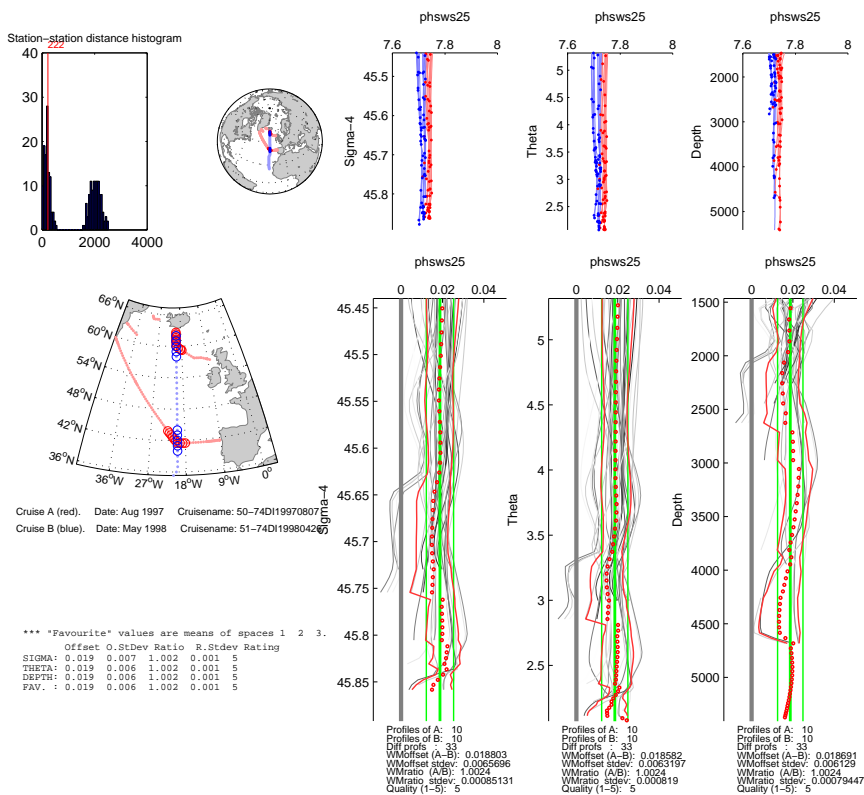


Figure 3. Crossover made with `cnaX` routines.

The application of various software packages (Tanhua et al., 2010) generated statistical and objective information about the differences between pairs of cruises, as well as the graphics needed to visually verify the computer determined differences. Each crossover analysis established the difference (i.e. the offset) between station profiles of two cruises that were located close to each other (2 degrees of Latitude was the typical distance). Only samples deeper than 1500 m were considered. The code returned the value of the offset as well as its uncertainty, and the number of contributing stations and samples.

In this work, manual, running-cluster and cnaX crossover procedures were applied on all possible pairs of CARINA cruises. Next, the crossover results were visually inspected in

order to ensure quality. Only “good” quality crossovers were selected, and those results were used for subsequent cruise correction calculations. Good crossovers were the ones with enough samples to produce reasonably uniform data profiles over the entire area included in the crossover. As additive offsets were being used for pH analysis, care was taken to verify that the profiles of both cruises flowed in parallel, so that the offset could be determined. Uncertainty as standard deviation was also used to provide more information about crossover quality.

In Figs. 2 and 3, two examples of pH crossovers are shown; Fig. 2 was produced by Running Cluster routines, and Fig. 3 was produced by cnaX routines (Tanhua et al., 2010).

4.2 Inversions (cruise corrections)

A weighted damped least squares inversion procedure (WDLSSQ) was applied to the collection of accepted crossover offsets and their uncertainties, to get the the solution of cruise corrections that would minimize the offsets (Gouretski and Jancke, 2001; Johnson et al., 2001). Uncertainties were used as weighting factors or data covariance matrix in this procedure. In order to ensure the highest quality results from the inversion and to get a more accurate and consistent solution, a small subset of cruises were a priori defined as “core”, giving to them a weighting factor of 2 in the inversion. These were chosen according to their geographical extent (i.e. covering a large distance) and expected high data quality (i.e. WOCE/CLIVAR quality), and were agreed upon by the CARINA Atlantic group. Offsets that involve “core” cruises received a higher weighting (double if only one cruise is “core”, and four times if both are “core”) in the inversion (Tanhua et al., 2010).

A second round of quality control was carried out by calculating pH from C_T and A_T for cruises with no pH measured, and including them into the crossover and inversion procedures. This was done to improve the confidence of corrections for cruises without so many crossovers, as pH data are sparse in some areas. An accuracy of ± 0.006 in calculated pH is accomplished from typical accuracies in on-boat measurement analysis of $\pm 3\text{--}4\text{ }\mu\text{mol kg}^{-1}$ in A_T and $\pm 2\text{ }\mu\text{mol kg}^{-1}$ in C_T (Millero, 2007; Zeebe and Wolf-Gladrow, 2001). The expected error in CARINA dataset is about $\pm 6\text{ }\mu\text{mol kg}^{-1}$ in A_T and $\pm 4\text{ }\mu\text{mol kg}^{-1}$ in C_T , so calculated pH accuracy could be lower than ± 0.006 here.

A total of 217 crossovers were used as input for the inversion procedure when using only measured pH data, and 311 when using also calculated pH from A_T and C_T .

Figure 4 shows the pH offsets for all individual crossovers before any adjustment (red dots), and after applying the full inversion solution to all cruises (i.e. all the corrections). As can be shown, after applying the corrections, the offsets are lower, and most of them fit between the ± 0.005 pH units boundary.

4.3 Adjustments

A minimum threshold of 0.005 pH units was defined and only corrections greater than this value has been taken into account by the CARINA group to decide the final adjustments for the cruises.

Corrections proposed by WDLSSQ inversion of running-cluster, cnaX and manual crossover offsets have been analyzed by the CARINA group to get an agreement upon the value of adjustment proposed and applied to the cruises.

4.4 Quality control

Once the adjustments were determined and applied for all analyzed parameters, a final crossover and inversion analysis

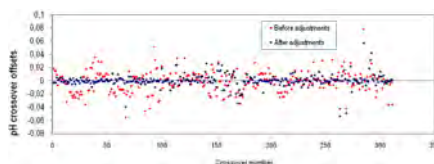


Figure 4. Crossover pH offsets obtained with original database and after adjustments were applied.

was performed for all cruises in CARINA database. Some additional regression analyses and statistical checks were also done in order to ensure the consistency of cruise data within their region and the internal consistency of carbon parameters. During this analysis, two Nordic Seas cruises, not previously adjusted by the crossover exercise because of too few data, evidently needed adjustments based on regional consistency and internal carbon consistency, so adjustments were calculated and applied to these data.

5 Assessment of applied adjustments

In this section, an assessment and description of the adjustments applied to cruises for CARINA database is made. CARINA identifiers for the cruises are the numbers inside the parentheses (see below). Exact data locations can be found at the CARINA website: http://cdiac.ornl.gov/oceans/CARINA/Carina_inv.html.

A set of figures and comments are presented for each cruise summarizing all crossover offsets and their standard deviation. Each figure shows the following information:

- Green dots: “Offsets”. These values are the offsets taken directly from each selected crossover. The standard deviation is shown as error bars on these dots. Mixed crossovers from running-cluster, cnaX, and manual crossovers have been used here
- Yellow line indicates the additive correction for the cruise calculated by WDLSSQ inversion. Note that the correction and offsets are of opposite sign.
- Black stars indicate the correction calculated by WDLSSQ inversion for the other cruises that intersect this cruise.
- Blue squares: “Predicted offset” shows the calculated offset that would be obtained by applying all inversion corrections to the cruises.
- Red dots: These are the residuals between the “Offsets” (Green dots) and “Predicted Offsets” (Blue squares)
- c suffix in the upper x-axis labels stands for Core Cruises.

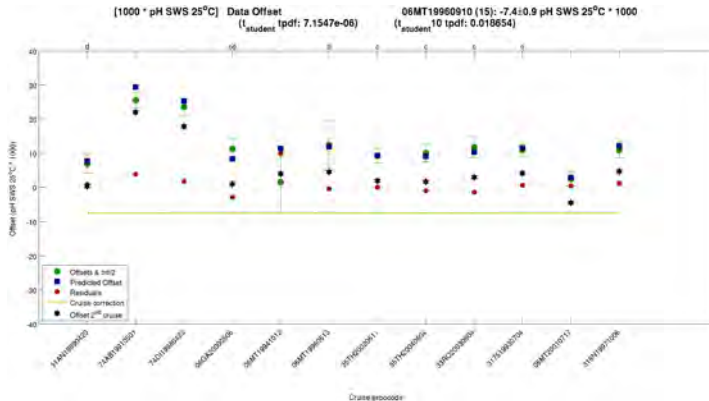


Figure 5. Cruise crossover information plot for 06MT19960910 (15).

- d suffix in the upper x-axis labels stands for cruises with pH derived from A_T and C_T .

5.1 Cruise 06MT19960910 (15) (Fig. 5)

This is the so called M36/5 cruise, carried out on board R/V Meteor, in the North Atlantic. It has 62 stations sampled with a 24 position rosette system. The cruise has 12 crossovers. The inversion suggests a correction of -0.007 ± 0.001 . Except one, all residuals are very low and fit inside ± 0.003 after the full solution of the inversion is applied. Very good fit exists with 5 core cruises. Based on this evidence, an adjustment of -0.007 was applied to the pH data.

5.2 Cruise 06MT20010507 (23) (Fig. 6)

This is leg 1 of the experiment called SFB460 (M50/1), carried out on board R/V Meteor in the subpolar North Atlantic. It has 53 stations sampled with a 24 position rosette system. The analysis of pH was done using spectrophotometric method with precision of ± 0.002 . Original data were reported on the Total pH scale at 21°C . The cruise has 7 crossovers. The inversion suggests a correction of -0.008 ± 0.005 . Except for two crossovers, all residuals are very low and fit inside ± 0.005 after the full solution of the inversion is applied. Very good fit also exists with two core cruises. Based on this evidence, an adjustment of -0.008 was applied to the pH data.

5.3 Cruise 06MT20010717 (25) (Fig. 7)

This is leg 4 of the experiment called SFB460 (M50/4), carried out on board R/V Meteor in the North Atlantic. It has

139 stations sampled with a 22 position rosette system. The analysis of pH was done using spectrophotometric method with precision of ± 0.002 and a standard deviation of replicates of ± 0.0009 , given an estimated uncertainty of ± 0.002 . Original data were reported on the Total pH scale at 21°C . The cruise has 16 crossovers. The inversion suggests a correction of -0.005 ± 0.001 . Except for two crossovers, all residuals are very low and fit inside ± 0.003 after the full solution of the inversion is applied. Very good fit exists with four core cruises and two GLODAP cruises (317519930704 and 06MT20030723). Based on this evidence, an adjustment of -0.005 was applied to the pH data.

5.4 Cruise 29CS19930510 (52) (Fig. 8)

This cruise is called MORENA-I, carried out on board R/V Cornide de Saavedra. It is a cruise along WOCE line AR16c. It has 92 stations sampled with a 24 position rosette system. Original data were reported on the NBS pH scale at 15°C . The cruise has 8 crossovers. The inversion suggests a correction of 0.017 ± 0.001 . All residuals are very low and fit inside ± 0.002 after the full solution of the inversion is applied. Very good fit also exists with three core cruises. Based on this evidence, an adjustment of 0.017 was applied to the pH data.

5.5 Cruise 29GD19821110 (53) (Fig. 9)

This is the so called GALICIA-V cruise, carried out on board R/V Garcia del Cid on Atlantic close to NW of Spain. It has 19 stations sampled on hydrocasts with 5L Niskin bottles. The analysis of pH was done using potentiometric method

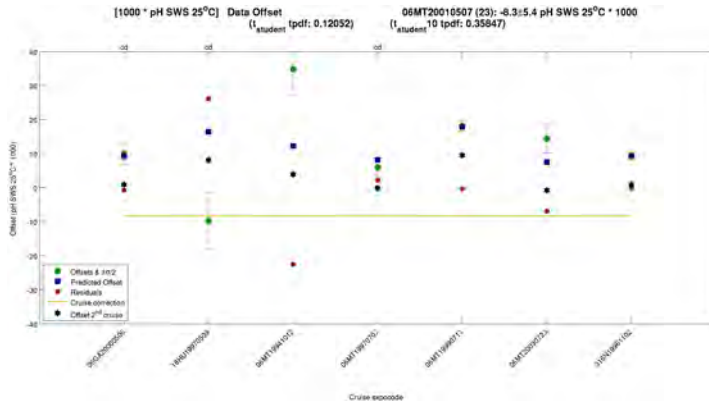


Figure 6. Cruise crossover information plot for 06MT20010507 (23).

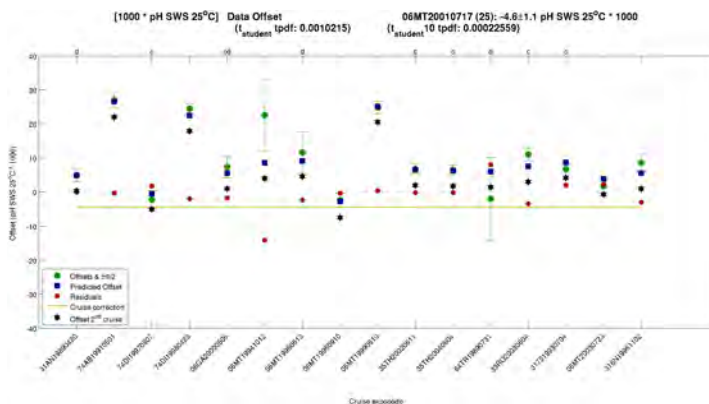


Figure 7. Cruise crossover information plot for 06MT20010717 (25).

with a glass electrode. The claimed accuracy is 0.003. The cruise has 7 crossovers. The inversion suggests a correction of 0.024 ± 0.002 . Except one, all residuals are low and fit inside ± 0.005 after the full solution of the inversion is applied. Good fit exits with 3 core cruises. Based on this evidence, an adjustment of 0.024 was applied to the pH data.

5.6 Cruise 29GD19840218 (55) (Fig. 10)

This is the so called GALICIA-VII cruise, carried out on board R/V Garcia del Cid on Atlantic close to NW of Spain. It has 33 stations sampled on hydrocasts with 1.7L Niskin bottles. The analysis of pH was done using potentiometric method with a glass electrode. The claimed accuracy is 0.003. The cruise has 8 crossovers. The inversion suggests a correction of 0.023 ± 0.001 . All residuals are low and fit inside ± 0.002 after the full solution of the inversion is applied.

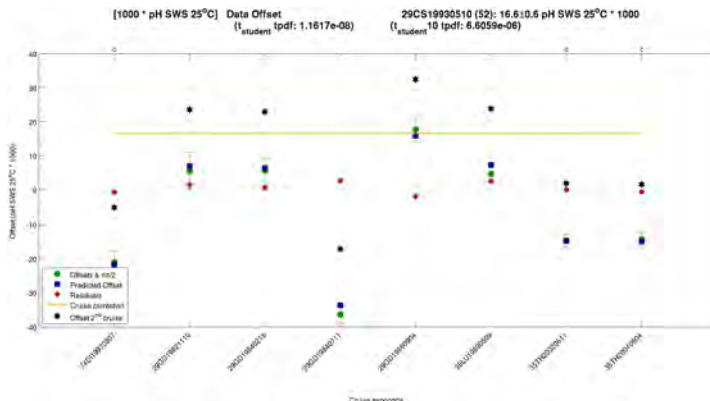


Figure 8. Cruise crossover information plot for 29CS19930510 (52).

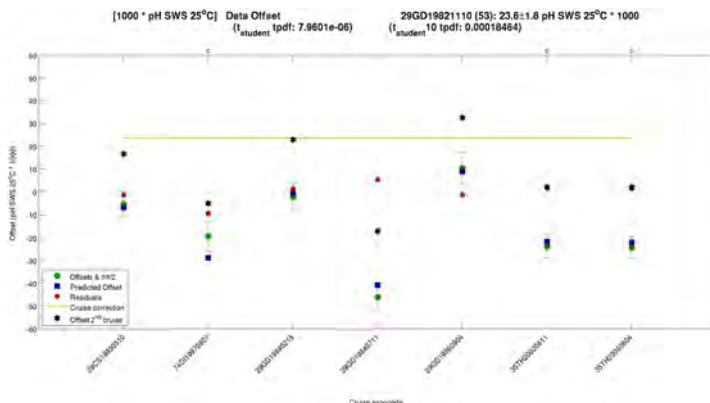


Figure 9. Cruise crossover information plot for 29GD19821110 (53).

Good fit exits with 3 core cruises. Based on this evidence, an adjustment of 0.023 was applied to the pH data.

5.7 Cruise 29GD19840711 (56) (Fig. 11)

This is the so called GALICIA-VIII cruise, carried out on board R/V Garcia del Cid on Atlantic close to NW of Spain. It has 118 stations sampled on hydrocasts with 1.7L Niskin bottles. The analysis of pH was done using potentiometric method with a glass electrode. The claimed accuracy is

0.003. The cruise has 8 crossovers. The inversion suggests a correction of -0.017 ± 0.002 . Except one, all residuals are low and fit inside ± 0.005 after the full solution of the inversion is applied. Good fit also exists with 3 core cruises. Based on this evidence, an adjustment of -0.017 was applied to the pH data.

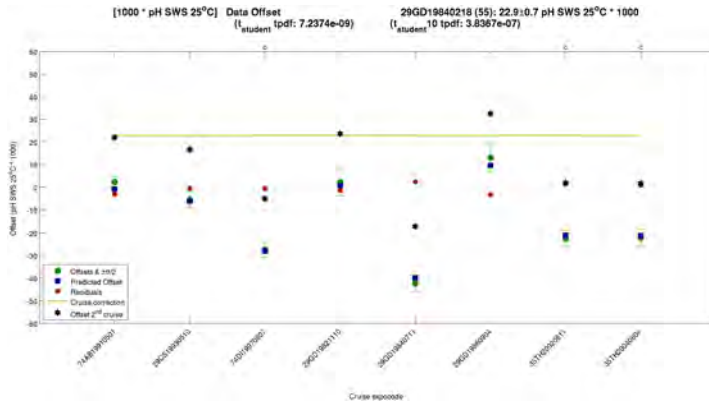


Figure 10. Cruise crossover information plot for 29GD19840218 (55).

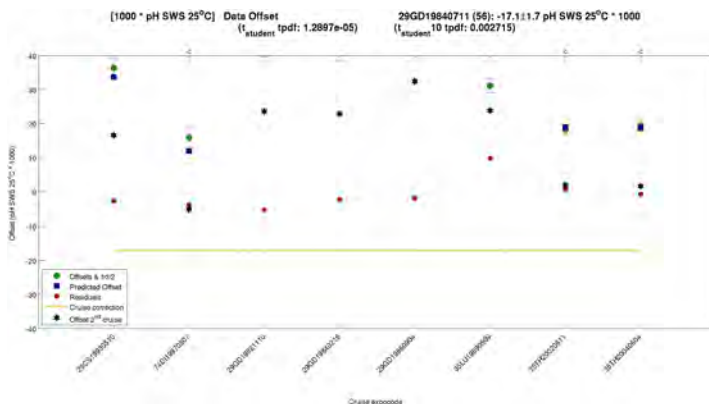


Figure 11. Cruise crossover information plot for 29GD19840711 (56).

5.8 Cruise 29GD19860904 (57) (Fig. 12)

This is the so called GALICIA-IX cruise, carried out on board R/V Garcia del Cid on Atlantic close to NW of Spain. It has 50 stations sampled on hydrocasts with 1.7L Niskin bottles. The analysis of pH was done using potentiometric method with a glass electrode. The claimed accuracy is 0.003. The cruise has 7 crossovers. The inversion suggests a correction of 0.032 ± 0.001 . All residuals are very low and fit inside ± 0.005 after the full solution of the inversion is ap-

plied. Very good fit also exists with 3 core cruises. Based on this evidence, an adjustment of 0.032 was applied to the pH data.

5.9 Cruise 29HE20010305 (61) (Fig. 13)

This is the so called FICARAM II cruise, carried out on board R/V Hesperides along WOCE section A17 in the western South Atlantic. It has 29 full depth stations sampled with a 24 position rosette system. The analysis of pH was

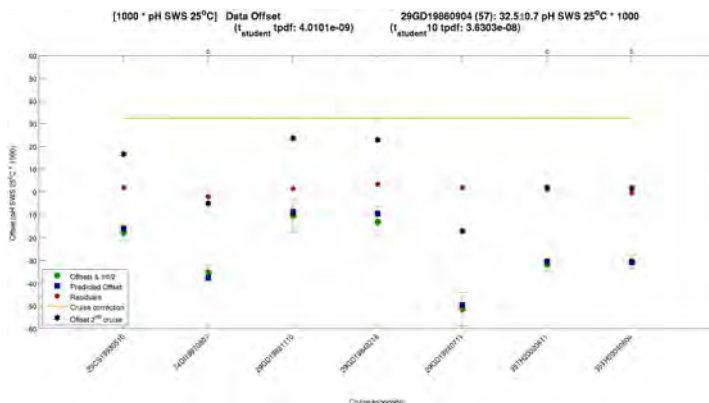


Figure 12. Cruise crossover information plot for 29GD19860904 (57).

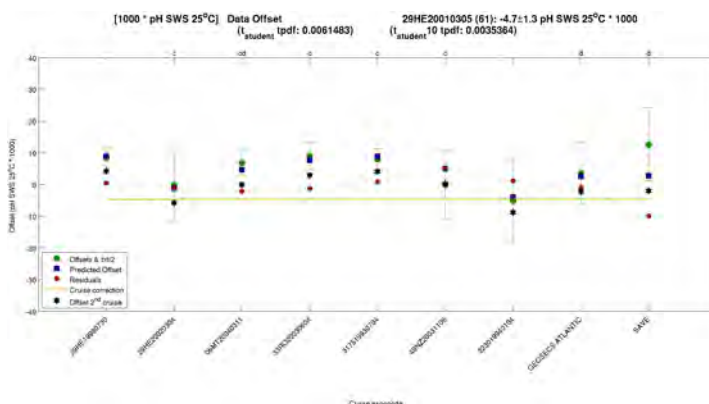


Figure 13. Cruise crossover information plot for 29HE20010305 (61).

done using spectrophotometric method. CRM batch 41 and 51 were used, with an uncertainty of 0.002. Original data were reported on the Total pH scale at 25 °C. The cruise has 9 crossovers. The inversion suggests a correction of $+0.005 \pm 0.001$. Except one, all residuals are low and fit inside ± 0.005 after the full solution of the inversion is applied. Very good fit also exists with 5 core cruises. Based on this evidence, an adjustment of $+0.005$ was applied to the pH data.

5.10 Cruise 33LK19960415 (84) (Fig. 14)

This is the so called ETAMBOT2 cruise, carried out on board R/V Edwin Link along WOCE section AR04h, in the west equatorial Atlantic, near Brazil. It has 94 stations sampled with a 24 position rosette system. The analysis of pH was done using the potentiometric method with a standard deviation of replicates of 0.003. Original data were reported on the Total pH scale at in-situ conditions of temperature and pressure. Data were measured at 25 °C.

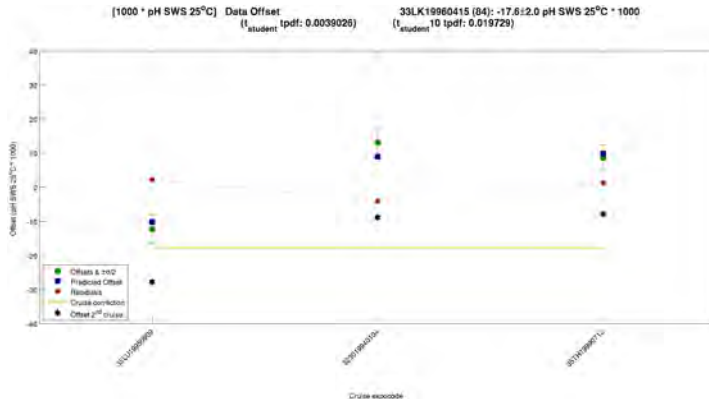


Figure 14. Cruise crossover information plot for 33LK19960415 (84).

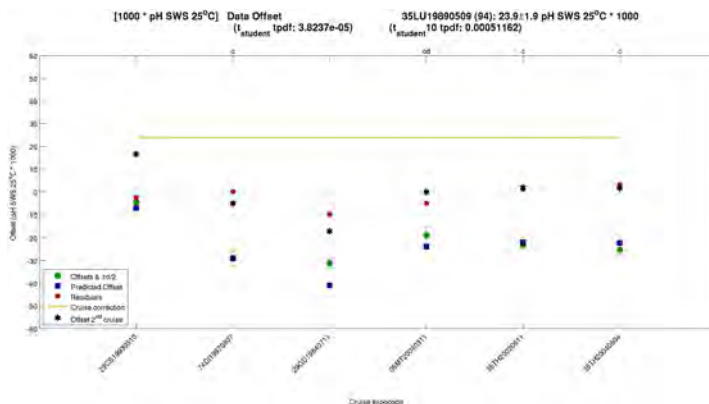


Figure 15. Cruise crossover information plot for 35LU19890509 (94).

The cruise has 3 crossovers. The inversion suggests a correction of -0.018 ± 0.003 . All residuals are very low and fit inside ± 0.005 after the full solution of the inversion is applied. Good fit also exists with one GLODAP cruise (323019940104). Based on this evidence, an adjustment of -0.018 was applied to the pH data.

5.11 Cruise 35LU19890509 (94) (Fig. 15)

This is the so called BORDEST-3 cruise carried out on board R/V “Le Noroit”, on a rectangular grid in the Atlantic west of the Iberian Peninsula. It has 47 full depth stations sampled with a rosette system. The cruise has 6 crossovers. The inversion suggests a correction of 0.024 ± 0.002 . Very good fit also exists with 4 core cruises. Based on this evidence, an adjustment of 0.024 was applied to the pH data.



Figure 16. Cruise crossover information plot for 35LU19950909 (95).

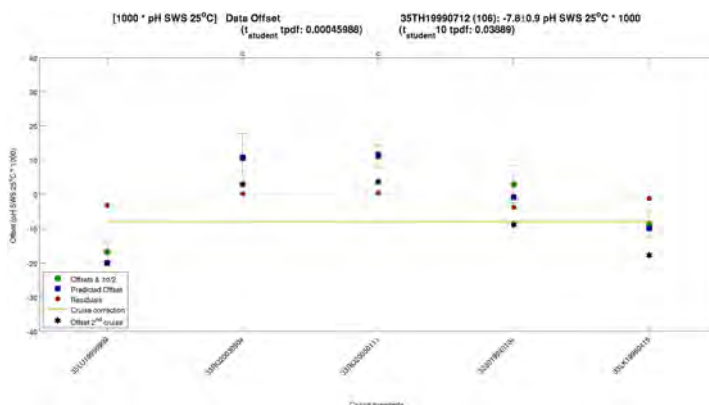


Figure 17. Cruise crossover information plot for 35TH19990712 (106).

5.12 Cruise 35LU19950909 (95) (Fig. 16)

This is the so called ETAMBOT1 cruise, carried out on board R/V “Le Noroit” along WOCE section AR04g, in the western equatorial Atlantic, near Brazil. It has 85 full depth stations sampled with a 24 position rosette system. The analyses of pH were done using the potentiometric method with a standard deviation of replicates of 0.002. Original data were reported on the Total pH scale at in-situ conditions for temperature and pressure. Data were measured at 25 °C. The

cruise has 3 crossovers. The inversion suggests a correction of -0.028 ± 0.003 . All residuals are very low and fit inside ± 0.005 after the full solution of the inversion is applied. Based on this evidence, an adjustment of -0.028 was applied to the pH data.

5.13 Cruise 35TH19990712 (106) (Fig. 17)

This is the so called EQUALANT99 cruise, carried out on board R/V “Thalassa” in the equatorial Atlantic. It has 102

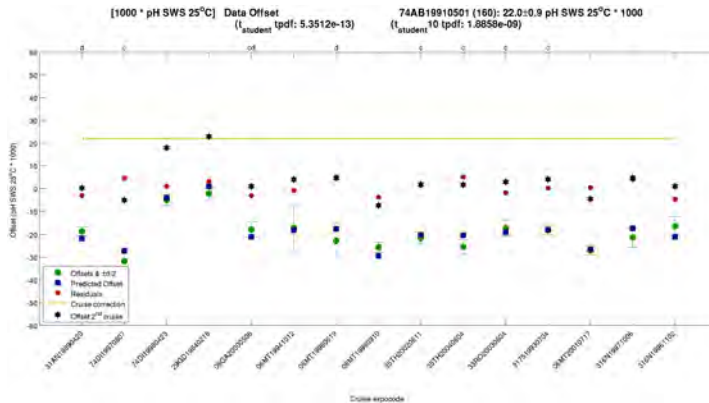


Figure 18. Cruise crossover information plot for 74AB19910501 (160).

stations sampled with a 24 position rosette system. The analysis of pH was done using the potentiometric method with a standard deviation of replicates of 0.003. Original data were reported on the Total pH scale at in-situ conditions of temperature and pressure. Samples were measured at 25 °C. The cruise has 5 crossovers. The inversion suggests a correction of -0.008 ± 0.001 . All residuals are very low and fit inside ± 0.003 after the full solution of the inversion is applied. Very good fit also exists with two GLODAP cruises (33RO20030604 and 33RO20030111). Based on this evidence, an adjustment of -0.008 was applied to the pH data.

5.14 Cruise 58AA19950217 (119)

This is the so called 58AA9502 cruise carried out on board R/V Haakon Mosby in the Nordic Seas. It has 34 stations sampled with a 12 position rosette system. The analysis of pH was done using spectrophotometric method, with a reported accuracy of ± 0.002 and a precision of 0.001. The analysis of the data showed that the scatter on pH is high. The cruise has only a few pH data, so decision here was to flag the measured pH data as questionable, and not include them in the data product.

5.15 Cruise 58JH19970414 (141)

This is the so called 58JH9704 cruise, carried out on board R/V "Johan Hjort" in the Nordic Seas. It has 135 stations sampled with a 12 position rosette system. The analysis of pH was done using spectrophotometric method. Original data were reported on the Total pH scale at 15 °C. Not enough crossovers with other cruises were found to support

an adjustment, but a comparison with Greenland Sea deep waters suggested a need of a correction of $+0.025$. Comparison with pH calculated from adjusted A_T and C_T supports this adjustment, and so does the MLR analysis (see below). Based on this evidence, an adjustment of 0.025 was applied to the pH data.

5.16 Cruise 58JH19980801 (142)

This is the so called 58JH9808 cruise, carried out on board R/V "Johan Hjort" in the Nordic Seas. It has 49 stations sampled with a 12 position rosette system. The analysis of pH was done using spectrophotometric method, with a reported precision of about ~ 0.005 pH units. Original data were reported on the Total pH scale at 15 °C. Not enough crossovers with other cruises were found to support an adjustment, but a comparison with Greenland Sea deep waters suggested a need of a correction of $+0.020$. Intercomparison with pH calculated from adjusted A_T and C_T supports this adjustment, and so does the MLR analysis (see below). Based on this evidence, an adjustment of 0.020 was applied to the pH data.

5.17 Cruise 74AB19910501 (160) (Fig. 18)

This cruise is the so called Vivaldi expedition, carried out on board R/V Charles Darwin. It has 614 stations, from which only 34 are deep stations. The stations were sampled with a 24 position rosette system. Original data were reported on the NBS pH scale at 15 °C. This cruise has 15 crossovers, and the inversion suggests a correction of 0.022 ± 0.001 . All residuals after the full solution of the inversion have been applied are very low and fit

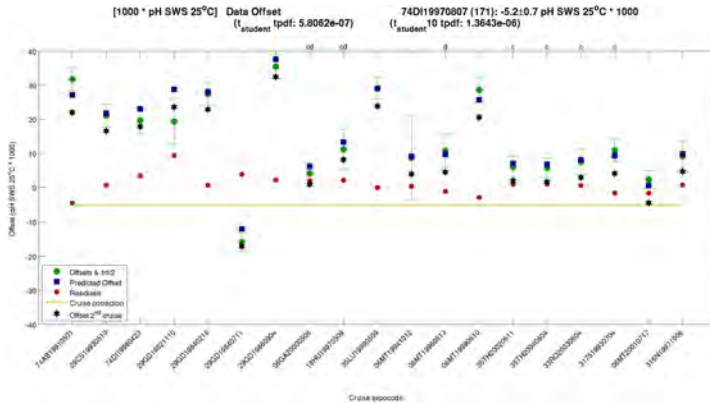


Figure 19. Cruise crossover information plot for 74DI19970807 (171).

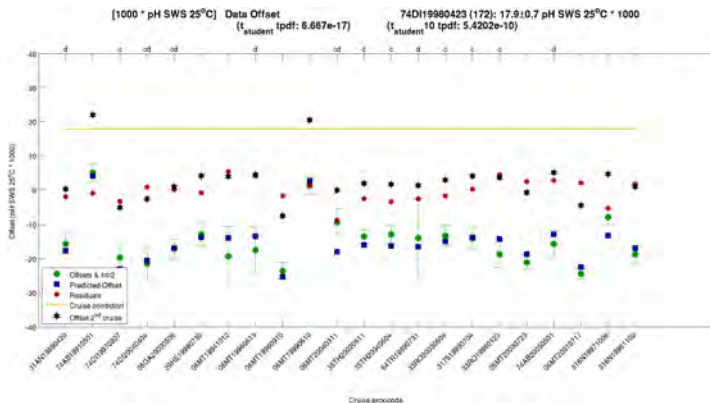


Figure 20. Cruise crossover information plot for 74DI19980423 (172).

inside ± 0.005 . Very good fit also exists with 5 GLODAP cruises (74DI19970807, 06MT19941012, 317519930704, 316N19971006 and 316N19961102). Based on this evidence, an adjustment of 0.022 was applied to the pH data.

5.18 Cruise 74DI19970807 (171) (Fig. 19)

This is the so called FOUREX cruise (IGY section Four Repeat Experiment), carried out on board R/V Discovery, along the WOCE leg A25. It has 143 full depth stations. The cruise

has 19 crossovers. The inversion suggests a correction of -0.005 ± 0.001 . Except one, all residuals are low and fit inside ± 0.005 after the full solution of the inversion is applied. Very good fit also exists with 6 core cruises. Based on this evidence, an adjustment of -0.005 was applied to the pH data.

5.19 Cruise 74DI19980423 (172) (Fig. 20)

This is a cruise on a meridional section along 20° W from 20° N to 65° N, carried out on board R/V Discovery. It has



Figure 21. Cruise crossover information plot for 91AA19971204 (183).

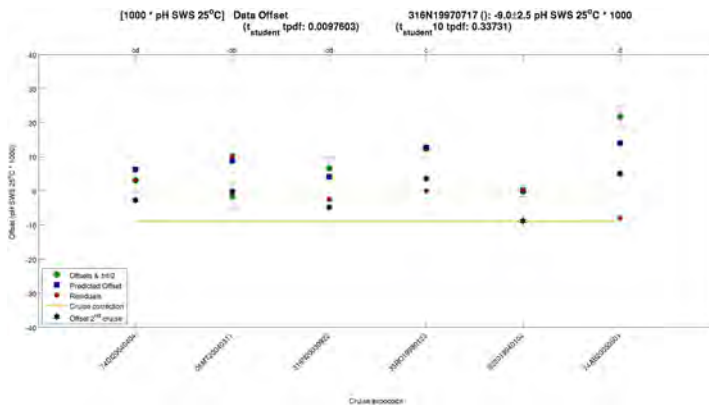


Figure 22. Cruise crossover information plot for 316N19970717.

44 full depth stations sampled with a 24 position rosette system. The cruise has 22 crossovers. The inversion suggests a correction of 0.018 ± 0.001 . Except two, all residuals are low and fit inside ± 0.005 after the full solution of the inversion is applied. Good fit also exists with 8 core cruises. Based on this evidence, an adjustment of 0.018 was applied to the pH data.

5.20 Cruise 91AA19971204 (183) (Fig. 21)

This is the so called SWEDARP 1997 expedition, carried out on board S.A Agulhas on a meridional section along 6°E , with 40 stations. The analysis of pH was done using spectrophotometric method. Original data were reported on the Total pH scale at 15°C . The cruise has only 2 crossovers. The inversion suggests a correction of 0.021 ± 0.005 . All residuals fit inside ± 0.010 after the full solution of the inversion is applied. Based on this evidence, an adjustment of 0.021 was applied to the pH data.

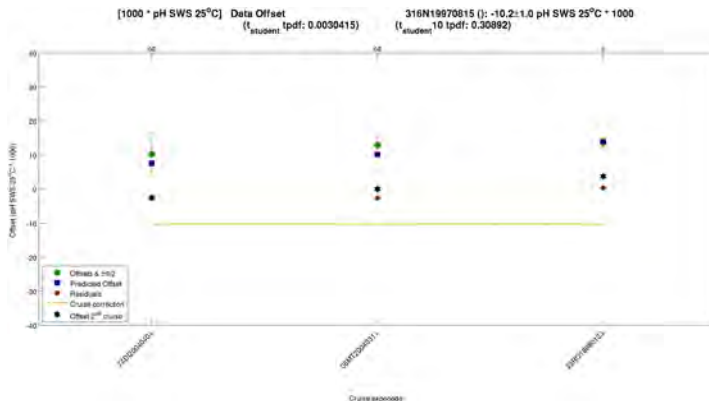


Figure 23. Cruise crossover information plot for 316N19970815.

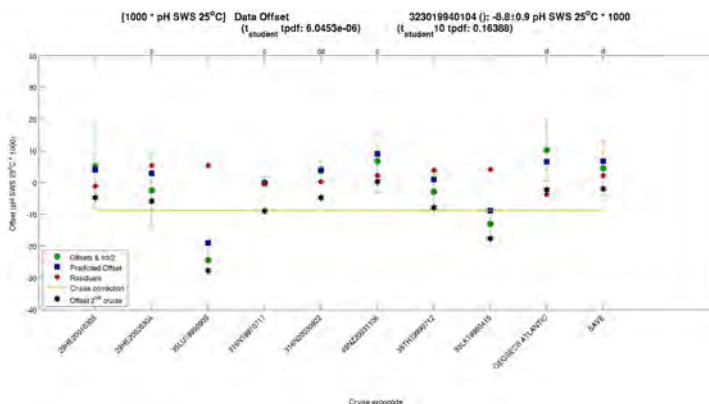


Figure 24. Cruise crossover information plot for 323019940104.

5.21 Cruise 316N19970717 (Fig. 22)

This is a cruise carried out on board R/V Knorr along WOCE section A20 in the North Atlantic. It has 95 stations sampled with a 36 position rosette system. The analysis of pH was done using the potentiometric method. CRM batch 33, 36 and 37 were used. The cruise has 6 crossovers. The inversion suggests a correction of -0.009 ± 0.003 . Except two, all residuals are very low and fit inside ± 0.005 after the full solution of the inversion is applied. Very good fit also exists with

two GLODAP cruises (32019940104 and 33RO19980123). Based on this evidence, an adjustment of -0.009 was applied to the pH data.

5.22 Cruise 316N19970815 (Fig. 23)

This is a cruise carried out on board R/V Knorr along WOCE section A22 in the North Atlantic. It has 7 stations sampled with a 36 position rosette system. The analysis of pH was done using the potentiometric method. CRM batch 33, 36



Figure 25. Cruise crossover information plot for 90MS19811009.

and 37 were used. The cruise has 3 crossovers. The inversion suggests a correction of -0.010 ± 0.001 . Very good fit also exists with two core cruises and one GLODAP cruise (33RO19980123). Based on this evidence, an adjustment of -0.010 was applied to the pH data.

5.23 Cruise 323019940104 (Fig. 24)

This is the so called CITHER 2 cruise carried out on board R/V Maurice Ewing, along WOCE section A17 in the western South Atlantic. It has 235 stations sampled with a 32 position rosette system. The analysis of pH was done using potentiometric method with stated overall precision of ± 0.003 . Original data were reported on the NBS scale at 15°C . The cruise has 10 crossovers. The inversion suggests a correction of -0.009 ± 0.001 . All residuals are low and fit inside ± 0.005 after the full solution of the inversion is applied. Very good fit also exists with 3 GLODAP cruises (316N1997017, GEOSECS.ATLANTIC and SAVE). Based on this evidence, an adjustment of -0.009 was applied to the pH data.

5.24 Cruise 90MS19811009 (Fig. 25)

This is the so called Weddell Polynya Expedition 81 (WE-POLEX 81) cruise carried out on board R/V Mikhail Somov in the Wedell Sea. It has 24 stations sampled with a 12 position rosette system. The analysis of pH was done using potentiometric method. The cruise has only 2 crossovers. The inversion suggests a correction of -0.034 ± 0.001 . All residuals are very low and fit inside ± 0.002 after the full solution of the inversion is applied. Based on this evidence, an adjustment of -0.034 was applied to the pH data.

Table 4. Adjustments applied to individual cruise files. “P” or “S” are acronyms for pH measurement method, and refer to Potentiometric or Spectrophotometric techniques respectively.

Cruise ID	Cruise expocode	Is Core Cruise	pH measurement method	Adjustment Applied
15	06MT19960910		P	-0.007
23	06MT20010507		S	-0.008
25	06MT20010717		S	-0.005
52	29CS19930510		P	0.017
53	29GD19821110		P	0.024
55	29GD19840218		P	0.023
56	29GD19840711		P	-0.017
57	29GD19860904		P	0.032
61	29HE20010305		S	0.005
84	33LK19960415		P	-0.018
94	35LU19890509		P	0.024
95	35LU19950909		P	-0.028
106	35TH19990712		P	-0.008
119	58AA19950217		S	Flagged 3
141	58JH19970414		S	0.025
142	58JH19980801		S	0.02
160	74AB19910501		P	0.022
171	74DI19970807	Y	S	-0.005
172	74DI19980423	Y	S	0.018
183	91AA19971204		S	0.021
-	323019940104	Y	P	-0.009
-	90MS19811009		P	-0.034
-	316N19970717	Y	P	-0.009
-	316N19970815	Y	P	-0.01

6 Results

Results are summarized in Table 4. This table shows the cruises in the three CARINA datasets (ATL, AMS and SO) for which pH adjustments have been applied. The following data is presented in the table:

- Cruise ID: CARINA assigned identification number for the cruise.
- Cruise Expocode: String that identifies the cruise. It is composed of a country code (two numbers), vessel code (two characters or numbers) and the departure date in year, month, day format (YYYYMMDD)
- Indicator for cruises used as “Core Cruises” in the crossover analysis.
- pH analysis method: Potentiometric (P) or Spectrophotometric (S)
- Adjustment: Adjustments applied for the cruises in the data products. All adjustments are fully supported by the CARINA group and no adjustments smaller than 0.005 pH units were applied.

7 Data quality evaluation

7.1 Overall level of internal consistency

A new crossover analysis was performed for the final adjusted CARINA data product, and the resulted offsets were used to estimate the internal consistency of the pH data, Fig. 26. The weighted mean (WM) was calculated for pH by using the absolute value of the offset (D) of the L crossovers with the uncertainty (σ):

$$WM = \frac{\sum_{i=1}^L D(i)/\sigma(i)^2}{\sum_{i=1}^L 1/\sigma(i)^2} \quad (2)$$

Based on this analysis we have estimated the overall level of internal consistency of the CARINA pH data to 0.005.

7.2 Overall evaluation

To make an overall evaluation of the dataset quality, a Multi-Linear Regression (MLR) was performed with potential temperature (θ), salinity, latitude, apparent oxygen utilisation (AOU), nitrate, phosphate and silicate in order to remove as much natural variability as possible. All of these parameters were included knowing that interdependence exists between many of them. However, the goal here is not to generate a statistical model for pH, but rather a way of getting residuals with natural variability removed to a large extent (adjusted R-square for MLR with full pH adjusted dataset of 0.96), so that they can be used to test the data quality. This analysis

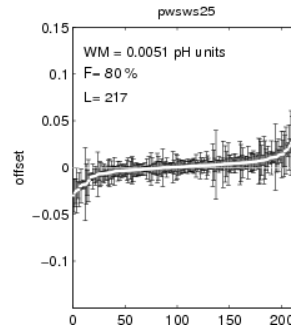


Figure 26. Sorted offsets calculated for the crossovers in the CARINA data after adjustments have been applied. WL: the weighted mean of the offsets (see text); F : the percentage of offsets indistinguishable from 0 within their uncertainty; L : the number of crossovers.

method facilitates a better assessment of scatter and biases in the pH dataset, and even to test the applied adjustments. A fact that has to be born in mind is that the MLR procedure can transmit the measurement error of the explanatory variables to the pH residuals, so this result has to be analyzed knowing its limitation.

In order to improve the quality of the evaluation, the MLR analysis was applied in four density layers. Density at 1000 db (σ_1) was used to divide the ocean in four layers. The upper thermocline was set by $\sigma_1 < 32.25 \text{ kg m}^{-3}$. Intermediate waters (depths from about 1000 to 2000 m) were defined by $32.25 \leq \sigma_1 < 32.39 \text{ kg m}^{-3}$. Waters between approximately 2000 to 3000 m were defined by $32.39 \leq \sigma_1 < 32.53 \text{ kg m}^{-3}$, which corresponds to North Atlantic Deep Waters (NADW). And finally, a fourth layer that applies for depths close to the bottom, where the presence of Antarctic Bottom Waters (AABW) dominates. This last layer was defined by $\sigma_1 > 32.53 \text{ kg m}^{-3}$. The surface layer with depths $< 200 \text{ m}$ was not used in this evaluation.

Using the MLR analysis, pH residuals were calculated by the following equation:

$$\text{pH}_{\text{MLR}} = \sum_{i=1}^8 a_i \cdot X_i \quad (3)$$

$$\text{pH}_{\text{residuals}} = \text{pH}_{\text{measured}} - \text{pH}_{\text{MLR}}$$

where X_i stand for Theta, Salinity, Latitude, AOU, Nitrate, Phosphate, Silicate, and a constant term. This procedure was done with CARINA corrected database, and also for the database without pH adjustments applied.

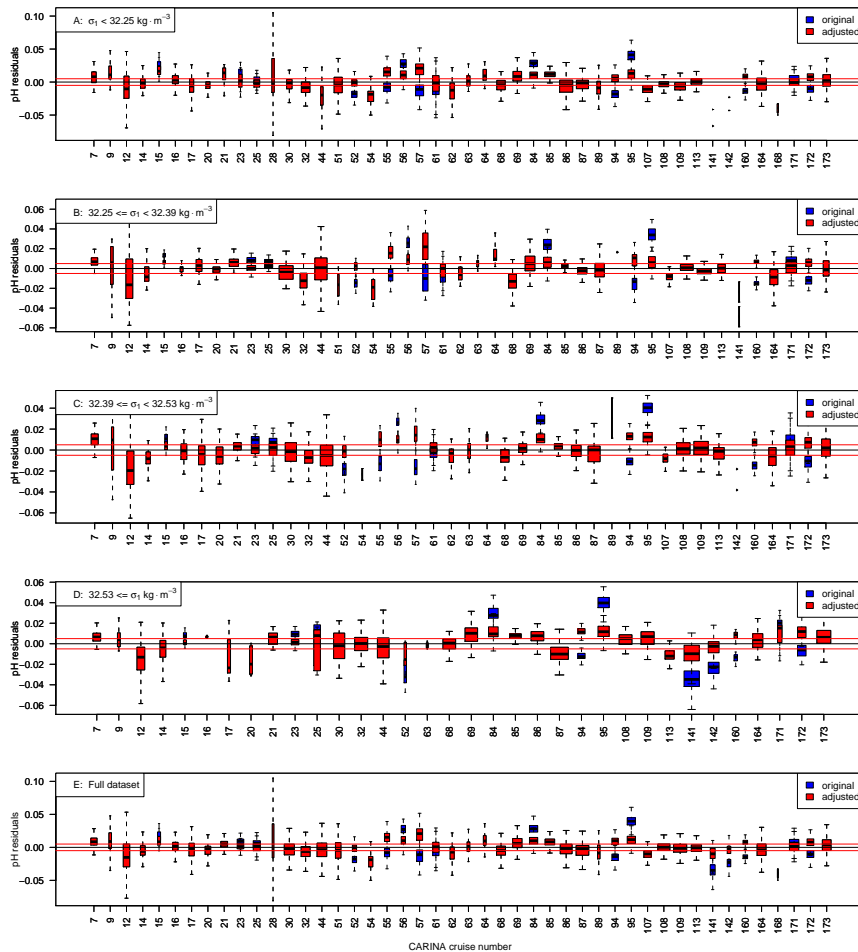


Figure 27. pH residuals obtained from the CARINA dataset by applying an MLR for pH data against Theta, Salinity, Latitude, AOU, Nitrate, Phosphate and Silicate. **(A)** to **(D)** are CARINA subsets for the indicated σ_1 intervals, and **(E)** is the combined residual for the full dataset. Blue values are residuals with the original unadjusted pH values, and red values are the final adjusted pH values. Red lines are the ± 0.005 pH units used as lower limit for adjustments in the secondary QC.

The pH residuals in each density layers are shown in the Fig. 27. The figure shows a box plot of pH residuals for each cruise. The box/whiskers represent the typical five number summary (minimum, 25th percentile, median, 75th percentile and maximum), the width of individual boxes corresponds to the number of samples for that cruise, and the numbers on the x-axis are the CARINA cruise IDs as referenced in Table 3.

The best MLR fit ($R^2 = 0.98$) is obtained for the shallower waters, with a mean residual standard error of 0.015. The next two layers have a slightly lower mean residual standard error (0.012 for layer 2 and 3 with a R^2 of 0.87 and 0.69 respectively) and the deepest layer has a residual standard error of 0.016 (R^2 of 0.77). In terms of mean deviation of each cruise, the mean standard error of the medians of all cruises is 0.009 pH units for each of three deepest density layers. The lower panel in the figure stands for the combined pH residuals of the four density layers.

The pH residuals are lower when using the corrected database, in comparison with the uncorrected original ones. Most of the cruises have the pH residuals median inside of the ± 0.005 boundary. In addition to the pH measurement errors, there are two other sources for the pH residuals: the MLR being not able to explain all the real variability of pH; and the measurement errors of the predictor parameters.

Acknowledgements. This work has been performed and funded as part of the EU project CarboOcean (Project 511176). Additional support from the International Ocean Carbon Coordination Project IOCCP (Maria Hood) and the Hanse Institute for Advanced Study (HWK Delmenhorst, Germany) are gratefully acknowledged. Additional support provided as follows: for A. Velo, F. F. Perez, and M. de la Paz, PGIDIT05OXIC40203PM Xunta de Galicia and CTM200627116E/MAR MEC; X. Lin, NOAA grant NA08OAR4310820; R. M. Key, NOAA grant NA08OAR4320752 and NA08OAR4310820; T. Tanhua, EU FP6 CARBOOCEAN Integrated Project, Contract no. 511176. We want to thank the work of the editor M. Hoppema and the reviewers, whose contribution have been very important for the quality and consistency of the manuscript. We want also to thank the hard work and dedication of analysts and investigators that collect the data at the sea; without them, this work would not be possible. We gratefully acknowledge also those who have contributed their data to the CARINA project.

Edited by: M. Hoppema

References

- Bates, R. and Vijh, A.: Determination of pH: Theory and practice, *J. Electrochem. Soc.*, 120, 263C–263C, 1973.
- Caldeira, K. and Wickett, M.: Anthropogenic carbon and ocean pH, *Nature*, 425, 365–365, 2003.
- Clayton, T. D. and Byrne, R. H.: Spectrophotometric seawater pH measurements: total hydrogen ion concentration scale calibration of m-cresol purple and at-sea results, *Deep-Sea Res.*, 40, 2115–2129, 1993.
- Dickson, A. and Riley, J.: The estimation of acid dissociation constants in seawater media from potentiometric titrations with strong base. I. The ionic product of water (K_w), *Mar. Chem.*, 7, 89–99, 1979.
- Dickson, A.: pH scales and proton-transfer reactions in saline media such as sea water, *Geochim. Cosmochim. Acta*, 48, 2299–2308, 1984.
- Dickson, A. and Millero, F.: A comparison of the equilibrium constants for the dissociation of carbonic acid in seawater media, *Deep-Sea Res.*, 34, 1733–1743, 1987.
- Dickson, A.: Standard potential of the reaction: $\text{AgCl (s)} + 1/2\text{H}_2\text{(g)} = \text{Ag (s)} + \text{HCl (aq)}$, and the standard acidity constant of the ion HSO_4^- in synthetic sea water from 273.15 to 318.15 K, *J. Chem. Thermodyn.*, 22, 113–127, 1990.
- Dickson, A. G.: The measurement of sea water pH, *Mar. Chem.*, 44, 131–142, 1993.
- Dickson, A. G., Sabine, C. L., and Christian, J. R.: Guide to best practices for ocean CO_2 measurements, *PICES Special Publication*, 3, 191 pp., 2007.
- Friis, K., Körtzinger, A., and Wallace, D.: Spectrophotometric pH measurement in the ocean: Requirements, design, and testing of an autonomous charge-coupled device detector system, *Limnol. Oceanogr. Methods*, 2, 126–136, 2004.
- Gouretski, V. V. and Jancke, K.: Systematic errors as the cause for an apparent deep water property variability: global analysis of the WOCE and historical hydrographic data, *Prog. Oceanogr.*, 48, 337–402, 2001.
- Hansson, I.: A new set of pH scales and standard buffers for sea water, *Deep-Sea Res.*, 20, 471–491, 1973.
- Johnson, G., Robbins, P., and Hufford, G.: Systematic adjustments of hydrographic sections for internal consistency, *J. Atmos. Ocean. Tech.*, 18, 1234–1244, 2001.
- Key, R. M., Tanhua, T., Olsen, A., Hoppema, M., Jutterström, S., Schirnick, C., van Heuven, S., Kozyr, A., Lin, X., Velo, A., Wallace, D. W. R., and Mintrop, L.: The CARINA data synthesis project: introduction and overview, *Earth Syst. Sci. Data*, 2, 105–121, doi:10.5194/essd-2-105-2010, 2010.
- Lewis, E., Wallace, D., and Allison, L.: Program developed for CO_2 system calculations, ORNL/CDIAC-105, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, US Department of Energy, Oak Ridge, Tennessee, 1998.
- Mehrbach, C., Culbertson, C., Hawley, J., and Pytkowicz, R.: Measurement of the apparent dissociation constants of carbonic acid in seawater at atmospheric pressure, *Limnol. Oceanogr.*, 897–907, 1973.
- Millero, F., Zhang, J., Fioll, S., Sotolongo, S., Roy, R., Lee, K., and Mane, S.: The use of buffers to measure the pH of seawater, *Mar. Chem.*, 44, 143–152, 1993.

- Millero, F.: The marine inorganic carbon cycle, *Chemical Reviews-Columbus*, 107, 308–341, 2007.
- Morris, A. and Riley, J.: The bromide/chlorinity and sulphate/chlorinity ratio in sea water, *Deep-Sea Res.*, 13, 699–705, 1966.
- Olafsson, J., Olafsdottir, S. R., Benoit-Cattin, A., Danielsen, M., and Takahashi, T.: North Atlantic Ocean acidification from time series measurements, *IOP Conf. Ser. Earth Environ. Sci.*, 6, 462005, 2009.
- Orr, J., Fabry, V., Aumont, O., Bopp, L., Doney, S., Feely, R., Gnanadesikan, A., Gruber, N., Ishida, A., and Joos, F.: Anthropogenic ocean acidification over the twenty-first century and its impact on calcifying organisms, *Nature*, 437, 681–686, 2005.
- Pérez, F. and Fraga, F.: The pH measurements in seawater on NBS scale, *Mar. Chem.*, 21, 315–327, 1987.
- Raven, J., Caldeira, K., Elderfield, H., Hoegh-Guldberg, O., Liss, P., Riebesell, U., Shepherd, J., Turley, C., and Watson, A.: Ocean acidification due to increasing atmospheric carbon dioxide, *The Royal Society*, London, 2005.
- Riley, J.: The occurrence of anomalously high fluoride concentrations in the North Atlantic, *Deep-Sea Res.*, 12, 219–220, 1965.
- Sørensen, S.: Enzymstudien II. Über die Messung und die Bedeutung der Wasserstoffionen-Konzentration bei enzymatischen Prozessen, *Biochem. Z.*, 21, 131–199, 1909.
- Takahashi, T., Williams, R., and Bos, D.: Carbonate chemistry, *GEOSCS Pacific Expedition*, 3, 1973–1974, 1982.
- Tanhua, T., Olsen, A., Hoppema, M., Jutterström, S., Schirnack, C., Van Heuven, S., Velo, A., Lin, X., Kozyr, A., Alvarez, M., Bakker, D. C. E., Brown, P., Falck, E., Jeansson, E., Lo Monaco, C., Olafsson, J., Perez, F. F., Pierrot, D., Rios, A. F., Sabine, C. L., Schuster, U., Steinfeldt, R., Stendardo, I., Anderson, L. G., Bates, N. R., Bellerby, R. G. J., Blindheim, J., Bullister, J. L., Gruber, N., Ishii, M., Johannessen, T., Jones, E. P., Köhler, J., Körtzinger, A., Metzl, N., Murata, A., Musielewicz, S., Omar, A. M., Olsson, K. A., de la Paz, M., Pfeil, B., Rey, F., Rhein, M., Skjelvan, I., Tilbrook, B., Wanninkhof, R., Mintrop, L., Wallace, D. W. R., and Key, R. M.: CARINA DATA SYNTHESIS PROJECT, ORNL/CDIAC-157, NDP-091, Carbon Dioxide Information Analysis Center OAK RIDGE NATIONAL LABORATORY, Oak Ridge, Tennessee, 37831–6335, 2009.
- Tanhua, T., van Heuven, S., Key, R. M., Velo, A., Olsen, A., and Schirnack, C.: Quality control procedures and methods of the CARINA database, *Earth Syst. Sci. Data*, 2, 35–49, doi:10.5194/essd-2-35-2010, 2010.
- van Heuven, S., Pierrot, D., Lewis, E., and Wallace, D.: MATLAB Program developed for CO₂ system calculations, ORNL/CDIAC-105b, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, US Department of Energy, Oak Ridge, Tennessee, 2009.
- Wootton, J., Pfister, C., and Forester, J.: Dynamic patterns and ecological impacts of declining ocean pH in a high-resolution multi-year dataset, *Proceedings of the National Academy of Sciences*, 105, 18848, 2008.
- Zeebe, R. and Wolf-Gladrow, D.: CO₂ in seawater: equilibrium, kinetics, isotopes, Elsevier Science, 2001.

Capítulo II

**A multiparametric method of interpolation using
WOA05 applied to anthropogenic CO₂ in the Atlantic**

ADVANCES IN MARINE CHEMISTRY
J. Blasco and J.M. Forja (eds.)

SCIENTIA MARINA 74S1
December 2010, 21-32, Barcelona (Spain)
ISSN: 0214-8358
doi: 10.3989/scimar.2010.74s1021

A multiparametric method of interpolation using WOA05 applied to anthropogenic CO₂ in the Atlantic

ANTON VELO, MARCOS VÁZQUEZ-RODRÍGUEZ, XOSE A. PADÍN,
MIGUEL GILCOTO, AÍDA F. RÍOS and FIZ F. PÉREZ

Instituto de Investigaciones Mariñas (CSIC). Eduardo Cabello 6, 36208 Vigo, Spain.
E-mail: avelo@iim.csic.es

SUMMARY: This paper describes the development of a multiparametric interpolation method and its application to anthropogenic carbon (C_{ANT}) in the Atlantic, calculated by two estimation methods using the CARINA database. The multiparametric interpolation proposed uses potential temperature (θ), salinity, conservative 'NO' and 'PO' as conservative parameters for the gridding, and the World Ocean Atlas (WOA05) as a reference for the grid structure and the indicated parameters. We thus complement CARINA data with WOA05 database in an attempt to obtain better gridded values by keeping the physical-biogeochemical sea structures. The algorithms developed here also have the prerequisite of being simple and easy to implement. To test the improvements achieved, a comparison between the proposed multiparametric method and a pure spatial interpolation for an independent parameter (O_2) was made. As an application case study, C_{ANT} estimations by two methods (qC_T^s and $TrOCA$) were performed on the CARINA database and then gridded by both interpolation methods (spatial and multiparametric). Finally, a calculation of C_{ANT} inventories for the whole Atlantic Ocean was performed with the gridded values and using ETOPO2v2 as the sea bottom. Thus, the inventories were between 55.1 and 55.2 Pg-C with the qC_T^s method and between 57.9 and 57.6 Pg-C with the $TrOCA$ method.

Keywords: CARINA, WOA05, CO₂, interpolation, multiparametric, anthropogenic carbon, back calculation.

RESUMEN: UN MÉTODO MULTIPARAMÉTRICO DE INTERPOLACIÓN UTILIZANDO WOA05, APLICADO AL CO₂ ANTROPOGÉNICO EN EL ATLÁNTICO. – Este trabajo describe el desarrollo de un método de interpolación multiparamétrico, y su aplicación al carbono antropogénico C_{ANT} en el Atlántico, calculado por dos métodos de estimación sobre la base de datos de CARINA. La interpolación multiparamétrica propuesta utiliza temperatura potencial (θ), salinidad, 'NO' y 'PO' conservativo a modo de parámetros conservativos para el mallado, y el World Ocean Atlas (WOA05) como referencia tanto para la estructura de la malla, como para los parámetros indicados. De este modo, este trabajo complementa CARINA con la base de datos de WOA05, intentando obtener mejores valores interpolados por el hecho de mantener las estructuras físico-biogeoquímicas marinas. Además, los algoritmos desarrollados tienen el prerequisite de ser sencillos y fáciles de implementar. Para comprobar las mejoras conseguidas, se ha realizado una comparación de un parámetro independiente (O_2) entre el método multiparamétrico y una interpolación puramente espacial. A modo de estudio de un caso de aplicación, se han realizado estimaciones de C_{ANT} mediante dos métodos (qC_T^s and $TrOCA$) sobre la base de datos de CARINA, y posteriormente interpolado mediante ambos métodos de interpolación (espacial y multiparamétrica). Por último, se ha realizado un cálculo de los inventarios de C_{ANT} para el Océano Atlántico completo con los valores interpolados y utilizando ETOPO2v2 como fondo marino. De este modo los inventarios obtenidos fueron de entre 55.1 y 55.2 Pg-C con la aproximación qC_T^s , y entre 57.9 y 57.6 Pg-C con la aproximación $TrOCA$.

Palabras clave: CARINA, WOA05, CO₂, interpolación, multiparamétrico, carbono antropogénico, retrocálculo.

INTRODUCTION

This work began as a contribution to the CARINA (Carbon in the Atlantic Ocean) Project, with the aim of developing an interpolation algorithm that would

enhance the gridding in low coverage areas, but with the premise of being easy to apply. The algorithm would also help to build a large and comprehensive carbon system database for the Atlantic Ocean with the CARINA database. The development should con-

tribute to estimate the C_{ANT} inventory of this ocean by, as part of a future work, gridding the C_{ANT} that can be calculated with the available approximation methods. Finally, the interpolation method could be used in the CARINA project to help to get the gridded product from the other available CARINA parameters. The interpolation algorithms developed here pursue the goal of being simple and easy to implement. After the seminal work of Gandin (1965) introducing objective analysis to produce gridded maps of meteorological variables in a systematic manner, objective interpolation methods were transferred from meteorology to oceanography in the late 1970s (Bretherton *et al.*, 1976; Freeland and Gould, 1976; Jalickee and Hamilton, 1977). Today, objective analysis appears in standard oceanography texts such as Bennett (1992) and Emery and Thomson (2001). In fact, one of the databases used in the present study, the World Ocean Atlas (WOA05, see Material and Methods section), was processed with these data analysis techniques. The present study uses a multiparametric inverse distance algorithm that was applied to the CARINA data (see the Material and Methods section) and took the WOA05 objective interpolated data as a reference to calculate the multiparametric distances. This approach provides a simple interpolation algorithm that is easy to use and to quality assess.

The CARINA Project has fed its dataset only from cruises in which carbon parameters were measured, so data coverage is low in certain regions, such as the Southern Ocean. Within this context, an interpolation method based only on geographical distances might perform poorly precisely on these regions due to the sparseness. A possible way to alleviate this problem is to incorporate more information in the interpolation algorithm other than the spatial. Thus, the consideration of fields of conservative properties for different water masses would become a benefit in this regard, providing better fits to real distributions than those generated from a purely spatial distance-based method. One additional advantage of this procedure is that the artefacts that may appear in the water mass distributions derived from plain spatial interpolations could be avoided.

The CARINA dataset is not distributed over a structured uniform grid, but is rather composed mainly from dispersed CTD stations organized in transoceanic sections. In terms of recorded variables, the dataset compiles many biogeochemical parameters, including the ones needed for the calculation of C_{ANT} by different methods. In contrast, the WOA05 dataset is structured in a homogenous three dimensional grid with thermohaline and biogeochemical variables defined at the nodes of the grid, but it lacks many of the parameters needed for carbon calculations. Therefore, the generation of a multiparametric interpolation algorithm that combines the properties of the WOA05 and CARINA datasets appears to be the logical way to proceed. Applying this to C_{ANT} es-

timations should provide C_{ANT} interpolated data over the structured WOA05 grid, taking advantage of the common hydrographical information available in both datasets.

As a way to evaluate the results, two versions of the interpolation method were compared; one based only on spatial distances and one that uses the physical and biogeochemical tracers (hereafter referred to as the Water Mass Properties [WMP] interpolation method). The contrast of the individual behaviour of the two methods was carried out using dependent variables (interpolating variables included in the multiparametric distances) and one independent variable (oxygen, not included in the multiparametric distances).

As the quality tests of the WMP interpolation method yield positive results, a step forward was taken: interpolating anthropogenic carbon over the WOA05 grid. The major role played by the oceans in the global carbon cycle is incontrovertible, since they have the capacity to sequester 2.2 ± 0.4 Pg C per year, roughly a 25% of the total anthropogenic carbon (C_{ANT}) emitted to the atmosphere (8.0 ± 0.5 Pg yr⁻¹) (Canadell *et al.*, 2007). Most outstandingly, the Atlantic Ocean stores 38% of the oceanic anthropogenic carbon (Sabine *et al.*, 2004), though it represents 29% of the global ocean surface area. The particular dynamics of the Atlantic Ocean allows the formation of deep waters in the North Atlantic and this enhances the uptake fluxes and storage capacity of C_{ANT} of this basin. Recently detected processes triggered by decadal changes of global climate, such as the slowdown of the Meridional Overturning Circulation, seem to have contributed significantly to reducing the sink capacity of C_{ANT} in both the North Atlantic and the Southern Ocean (Joos *et al.*, 1999; Le Quéré *et al.*, 2007). The juxtaposition of these opposed effects has dramatically magnified the need to accurately estimate the state of C_{ANT} inventories and has raised the importance of fine-tuning the C_{ANT} interpolation methods applied to sparse or geographically disperse datasets.

Two methods were used to obtain the estimation of C_{ANT} over the whole Atlantic Ocean, φC_T and $TrOCA$. The φC_T estimation method was chosen as it was developed by the authors (Vázquez-Rodríguez *et al.*, 2009a) and it was straightforward to apply and verify (the MATLAB script is publicly available for download at <http://oceanio.iim.csic.es/co2group/>). It is also an updated method and seems to perform well in comparison with other methods (Vázquez-Rodríguez *et al.*, 2009b). The $TrOCA$ method (Touratier *et al.*, 2007) was additionally considered as a support reference due to its ease of application. Consequently, anthropogenic carbon was calculated by applying these estimations to the CARINA dataset, and then gridded by both the WMP and Spatial interpolation methods. The next step taken was to calculate the volumes in order to obtain the inventories. ETOPO2v2 (U.S. Department of Commerce, 2006) was chosen as reference for the ocean floor in these calculations.

MATERIALS AND METHODS

Database

CARINA is a database of comprehensive carbon data, sourced from hydrographic cruises conducted in the Arctic, Atlantic and Southern Oceans. The project was initiated in 1999 as an essentially informal and unfunded project in Kiel, Germany, with the main goal of creating a database of relevant carbon variables to be used for accurate assessments of carbon inventories, transports and uptake rates. The CARINA data have been gathered from various sources and then put under rigorous quality controls (QC) to produce a consistent data product. Experience with previous synthesis efforts like the Global Data Analysis Project (GLODAP) (Key *et al.*, 2004) demonstrated that a consistent data product can be achieved from different cruises, performed by different laboratories and in very different regions. The CARINA database includes data and metadata from 188 oceanographic cruises or projects, (Hoppema *et al.*, 2009; Tanhua *et al.*, 2009; Key *et al.*, 2010; Tanhua *et al.*, 2010). In addition, 52 WOCE/GLODAP cruises were included in the quality control to ensure consistency with historical data. Parameters included in the CARINA dataset are salinity (S), potential temperature (θ), oxygen (O_2), nitrate (NO_3), phosphate (PO_4), silicate (SiO_4), total alkalinity (A_T), fugacity of carbon dioxide (fCO_2), total inorganic carbon (C_T), pH, CFC-11, CFC-12, CFC-113 and CCl4. Due to the different origins of the data, the data density has heterogeneous distributions, being scarcer in the South Atlantic than in the North Atlantic.

The World Ocean Atlas 2005 (WOA05) has widely proven its usefulness to the oceanographic and atmospheric research communities. WOA05 offers a gridded database interpolated from many different sources by oceanographic objective analysis techniques. The WOA05 climatological analyses were carried out on a $1^\circ \times 1^\circ$ grid. This comes from the fact that higher resolution analyses are not justified for all the measured properties, and they should be analyzed in the same manner. For a description of the WOA05 data and statistical fields, refer to <http://www.nodc.noaa.gov/OC5/WOA05/pubwoa05.html>. The site includes a list of values and statistical data in a one-degree latitude-longitude world grid (360×180) at 33 standard depth levels from the surface to a maximum depth of 5500 m.

The WOA05 series include analysis of temperature (Locarnini *et al.*, 2006), salinity (Antonov *et al.*, 2006), dissolved oxygen, apparent oxygen utilization, oxygen saturation (Garcia *et al.*, 2006a), and dissolved inorganic nutrients (Garcia *et al.*, 2006b). The climatologies defined here come from historical oceanographic profiles and selected data at different depths. Data used in the WOA05 were analyzed in a consistent, objective analysis mode and interpolated over a one-degree latitude-longitude grid at standard depth levels.

The aim of the WOA05 maps is to illustrate the large-scale characteristics of the distribution of ocean temperature. The fields used to generate these climatological maps were computed by objective analysis of quality-controlled historical temperature data. Maps are presented for climatological composite periods (annual, seasonal, monthly, and monthly difference fields from the annual mean field, and the number of observations) at selected standard depths. The annual climatology was calculated using all data regardless of the month of the observation. Seasonal climatologies were calculated using only data from the defined season (regardless of year). However, in this study only the annual maps for the whole Atlantic Ocean (Lat: $90^\circ S$ - $90^\circ N$, Lon: $81^\circ W$ - $33^\circ E$) were used.

For the C_{ANT} inventory calculations, the ETOPO2v2 (USDC, NOAA, NGDC 2006) bathymetry was used as bottom reference to calculate the volume of the deepest boxes.

Interpolation Method

Hereafter var is the parameter that we wish to interpolate. The objective is to obtain for each node j of the structured WOA05 grid an average value of var , which is unavailable from this database. The known values of var from the CARINA unstructured-grid nodes i will be used to fill in the j WOA05 nodes. A classic interpolation scheme of inverse distance is applied, using a weighted estimation of the i neighbouring samples to the j WOA05 node:

$$var_j = \frac{\sum_i var_i (f_i^j)^{-1}}{\sum_i (f_i^j)^{-1}} \quad (1)$$

The weighting factors used in the basic spatial interpolation are defined by:

$$f_i^j = w_{lat} \left(\frac{lat^i - lat^j}{\Delta lat} \right)^2 + w_{lon} \left(\frac{lon^i - lon^j}{\Delta lon} \right)^2 + w_z \left(\frac{z^i - z^j}{\Delta z} \right)^2 + w_\theta \left(\frac{\theta^i - \theta^j}{std(\theta^i)} \right)^2 + w_S \left(\frac{S^i - S^j}{std(S^i)} \right)^2 + w_{NO} \left(\frac{NO^i - NO^j}{std(NO^i)} \right)^2 + w_{PO} \left(\frac{PO^i - PO^j}{std(PO^i)} \right)^2 \quad (2)$$

where several variables are pondered with arbitrary weights (w_x). The specific values of the w_x terms to produce the f_i^j factors are subjected to the criteria of the researcher. The information concerning the geographical position is taken into account in the interpolation through the spatial coordinates of longitude (lon), latitude (lat) and depth (z). The interpolation factors are also determined using the information from four tracers: salinity (S), potential temperature (θ), 'NO'

and 'PO'. Both 'NO' ($=9[\text{NO}_3]+[\text{O}_2]$) and 'PO' ($=135[\text{HPO}_4]+[\text{O}_2]$) are conservative parameters, and like *potential temperature* and *salinity* they are characteristic of each water mass (Broecker, 1974; Ríos *et al.*, 1989; Pérez *et al.*, 1993). The Δlat , Δlon and Δz appearing in Equation (2) are the intervals of latitude, longitude and depth. Three different intervals were taken as results of sample availability in the CARINA database. First, a $\pm 2^\circ \times \pm 2^\circ$ (Lat-Lon) window was used. A larger area of $\pm 10^\circ \times \pm 10^\circ$ (Lat-Lon) was chosen if fewer than 20 samples were found in the previous boundary, and if again no samples were found within this interval, then the boundaries were expanded to an even bigger window of $\pm 20^\circ \times \pm 20^\circ$ (Lat-Lon). For depth, an interval of $\pm(150 + 0.1 \times \text{depth})$ metres was used. The tracer (S , θ , NO and PO) differences between the WOA05 and CARINA nodes are normalized using the standard deviation from each tracer computed in the corresponding equivalent volume defined by the intervals of latitude, longitude and depth. The quotient terms of the spatial or tracer differences with their spatial intervals or tracer standard deviations are then squared to convert them into the classical inverse quadratic distance interpolation equation.

Two different kinds of interpolations were applied to produce three-dimensional O_2 and C_{ANT} fields. The first one is designated as the "spatial interpolation" and is defined from the following weights:

$$w_{\text{lat}}, w_{\text{lon}}, w_z, w_{\theta}, w_S, w_{\text{NO}}, w_{\text{PO}} = (1, 1, 1, 0, 0, 0, 0) \quad (3)$$

These factors avoid the influence of the tracer properties in the interpolation. On the other hand, the "WMP interpolation" stands for the interpolation without weights in the spatial coordinates (latitude, longitude and depth) and using only the tracer variables:

$$w_{\text{lat}}, w_{\text{lon}}, w_z, w_{\theta}, w_S, w_{\text{NO}}, w_{\text{PO}} = (0, 0, 0, 1, 1, 1, 1) \quad (4)$$

C_{ANT} estimation methods

Two recently developed C_{ANT} back-calculation methods, φC_T° and TrOCA , were selected to determine C_{ANT} in the present study. Both methods separate the contributions to total carbon (C_T) from organic matter remineralization and CO_3Ca dissolution in a similar mode. However, there are characteristic distinctions. The TrOCA method uses a constant Redfield ratio (R_c) value of 1.35 (Kortzinger *et al.*, 2001), while the φC_T° method, following the ΔC^* method, uses the constant R_c ratio (1.45) proposed by Anderson and Sarmiento (1994). The most important difference between the two methods, though, lies in the way the reference for C_{ANT} -free waters is obtained. The TrOCA method estimates C_{ANT} using the following simple relationship:

$$C_{\text{ANT}} = \frac{(\text{TrOCA} - \text{TrOCA}^0)}{a} \quad (5)$$

where TrOCA represents a quasi-conservative tracer calculated from O_2 , C_T and A_T as follows:

$$\text{TrOCA} = \text{O}_2 + a (C_T - 0.5 A_T) \quad (6)$$

The TrOCA^0 reference represents the TrOCA tracer without any anthropogenic carbon influence, i.e. the pre-industrial TrOCA :

$$\text{TrOCA}^0 = e^{b+c\theta+\frac{d}{A_T^2}} \quad (7)$$

The coefficients a , b , c and d in the above equations are properly defined and established in Touratier *et al.* (2007). The TrOCA^0 equation is obtained from $\Delta^{14}\text{C}$ and CFC-11 data in the global ocean. The $\Delta^{14}\text{C}$ data are used to establish which water parcels can be assumed to be free of C_{ANT} . When the concentration of $\Delta^{14}\text{C} < 175\text{‰}$, the age of the corresponding water mass is greater than 1400 years, long before the massive emissions of CO_2 by humans had begun. The samples with maximum CFC-11 concentrations, typically between 262.9 and 271.3 pptv and corresponding to surface waters in 1992-1995 (maximum atmospheric $p\text{CFC-11}$), were also selected as part of the dataset to obtain the TrOCA^0 expression. Touratier *et al.* (2007) estimated an uncertainty of $\pm 6.2 \mu\text{mol kg}^{-1}$ in C_{ANT} determination for the TrOCA method, using an error propagation technique as in numerous previous studies (Gruber *et al.*, 1996; Sabine *et al.*, 1999).

The φC_T° method (Vázquez-Rodríguez *et al.*, 2009a) shares similar fundamentals with the ΔC^* back-calculation method (Lee *et al.*, 2003). The sub-surface layer (100-200 m) is taken in the φC_T° method as a reference for characterizing water mass properties at the moment of their formation. The air-sea CO_2 disequilibrium (ΔC_{dis}) is parameterized at the sub-surface layer first using a short-cut method (Thomas and Ittekkot, 2001) to estimate C_{ANT} . Since the average age of the water masses in the 100-200 m depth domain, and most importantly in outcropping regions it is under 25 years, the use of the short-cut method to estimate C_{ANT} is appropriate (Matear *et al.*, 2003). The pre-industrial total alkalinity A_T^0 and ΔC_{dis} parameterizations (in terms of conservative tracers) obtained from sub-surface data are applied directly to calculate C_{ANT} in the water column for waters above the 5°C isotherm and via an Optimum MultiParameter analysis (OMP) for waters with $\theta < 5^\circ\text{C}$. This procedure especially improves the estimates in cold deep waters that are subject to strong and complex mixing processes between Arctic and Antarctic water masses. Waters below the 5°C isotherm also represent an enormous volume of the global ocean (~86%). One important aspect of the φC_T° method is that none of the A_T^0 or ΔC_{dis} parameterizations are CFC -reliant. In addition, the φC_T° method proposes an approximation to the temporal and spatial variability of ΔC_{dis} ($\Delta\Delta C_{\text{dis}}$) in the Atlantic Ocean in terms of C_{ANT} and

ΔC_{dis} itself. Also, the small increase in A_T^o since the Industrial Revolution due to CaCO_3 dissolution changes projected from models (Heinze, 2004), and the effect of rising sea surface temperatures on the parameterized A_T^o are accounted for in the parameterizations. These two last corrections are minor but should still be considered if one wished to avoid a maximum $4 \mu\text{mol kg}^{-1}$ bias ($2 \mu\text{mol kg}^{-1}$ on average) in C_{ANT} estimates. The φC_T^o method expression for the calculation of C_{ANT} is as follows:

$$C_{ANT} = \frac{\Delta C^* - \Delta C_{dis}^i}{1 + \varphi \frac{[\Delta C_{dis}^i]}{C_{ANT}^{sat}}} \quad (8)$$

The ΔC^* method is defined after Gruber *et al.* (1996) as:

$$\Delta C^* = C_T - \frac{AOU}{R_c} - 0.5(PA_T - PA_T^0) - C_T^{eq} \quad (9)$$

The constant term φ is a proportionality factor that stands for the $\Delta C_{dis}/\Delta C_{dis}^i$ ratio and its value (0.55) is properly discussed in Vázquez-Rodríguez *et al.* (2009a). The ΔC_{dis}^i and PA_T^o terms are parameterized as a function of conservative parameters exclusively (Vázquez-Rodríguez *et al.*, 2009a). The C_{ANT}^{sat} stands for the theoretical C_{ANT} saturation concentration depending on the $p\text{CO}_2$ at the time of water masses formation (WMF), and is defined as $C_{ANT}^{sat} = S/35 (0.85\theta + 46.0)$ (at present $x\text{CO}_2$ air). Based on earlier uncertainty and error evaluations (Gruber *et al.*, 1996; Sabine *et al.*, 1999; Lee *et al.*, 2003; Touratier *et al.*, 2007), an estimated overall uncertainty of $\pm 5.2 \mu\text{mol kg}^{-1}$ is obtained for the φC_T^o method. This is in agreement with the av-

TABLE 1. – Mean and standard deviation (STD) of the interpolated data minus WOA05 data are computed for the whole Atlantic, for specific sub domains and for the spatial and WMP interpolation. The determination coefficient (r^2) between WOA05 and interpolated fields is also provided. Temperature is in °C. Salinity in psu and NO , PO and O_2 in $\mu\text{mol kg}^{-1}$.

Total	r^2	Total Mean	Std	r^2	$\theta > 5^\circ\text{C}$ Mean	Std	r^2	$\theta < 5^\circ\text{C}$ Mean	Std
θ spatial	0.964	0.34	1.02	0.856	0.23	0.62	0.914	0.55	1.48
θ WMP	0.993	0.10	0.43	0.952	0.11	0.37	0.988	0.07	0.53
Sspatial	0.925	0.01	0.14	0.887	-0.02	0.07	0.905	0.04	0.22
SWMP	0.984	0.00	0.07	0.954	-0.01	0.05	0.981	0.00	0.09
NOspatial	0.948	0.8	18.8	0.845	3.8	15.5	0.918	-4.7	22.7
NOWMP	0.989	2.5	8.6	0.957	3.6	8.2	0.983	0.4	9.2
POspatial	0.956	-5.7	18.6	0.893	-4.3	15.3	0.919	-8.2	23.2
POWMP	0.991	-3.0	8.4	0.971	-3.1	8.1	0.984	-2.7	8.7
O_2 spatial	0.901	4.1	14.5	0.857	4.4	13.1	0.918	3.7	16.6
O_2 WMP	0.943	2.6	11.0	0.945	2.7	8.4	0.921	2.6	15.0
Lat > 30°N									
θ spatial	0.934	0.20	1.24	0.860	0.34	1.61	0.80	0.01	0.42
θ WMP	0.987	0.06	0.55	0.987	0.06	0.55	0.82	0.04	0.41
Sspatial	0.832	0.02	0.20	0.752	0.03	0.27	0.72	0.00	0.06
SWMP	0.948	0.01	0.11	0.948	0.01	0.11	0.87	0.01	0.09
NOspatial	0.954	-1.0	12.7	0.916	-3.6	15.7	0.64	2.3	6.3
NOWMP	0.988	0.7	6.6	0.988	0.7	6.6	0.82	2.4	4.3
POspatial	0.937	-5.1	15.6	0.877	-6.3	19.6	0.57	-3.6	8.2
POWMP	0.988	-2.2	7.0	0.988	-2.2	7.0	0.82	-1.6	4.8
O_2 spatial	0.928	0.6	9.3	0.875	0.1	11.2	0.88	1.1	6.1
O_2 WMP	0.946	-0.6	8.2	0.946	-0.6	8.2	0.91	0.5	5.4
Tropical									
θ spatial	0.982	0.30	0.83	0.958	0.59	1.11	0.985	0.03	0.14
θ WMP	0.998	0.06	0.27	0.995	0.11	0.38	0.994	0.02	0.08
Sspatial	0.956	0.01	0.12	0.943	0.03	0.17	0.981	-0.01	0.02
SWMP	0.995	0.00	0.04	0.994	0.00	0.06	0.993	0.00	0.01
NOspatial	0.981	1.6	11.0	0.968	-1.7	13.6	0.898	4.7	6.3
NOWMP	0.995	2.5	5.3	0.992	1.8	6.6	0.964	3.2	3.8
POspatial	0.979	-2.2	11.6	0.967	-5.1	13.4	0.858	0.6	8.6
POWMP	0.995	-1.0	5.7	0.992	-1.6	6.5	0.955	-0.5	4.8
O_2 spatial	0.926	5.7	14.1	0.83	8.0	19.0	0.970	3.5	5.8
O_2 WMP	0.937	5.7	12.8	0.86	6.6	17.2	0.963	4.8	6.5
Lat < 30°S									
θ spatial	0.92	0.44	1.05	0.82	0.77	1.93	0.74	0.38	0.75
θ WMP	0.983	0.14	0.47	0.97	-0.04	0.66	0.92	0.17	0.43
Sspatial	0.77	0.00	0.13	0.82	0.09	0.22	0.75	-0.02	0.09
SWMP	0.96	-0.01	0.05	0.97	-0.01	0.09	0.94	-0.01	0.05
NOspatial	0.82	0.9	24.9	0.82	-13.5	39.9	0.53	3.7	19.5
NOWMP	0.96	3.4	11.1	0.96	-1.2	15.5	0.86	4.0	10.1
POspatial	0.85	-8.7	23.2	0.83	-18.2	38.5	0.65	-6.8	18.2
POWMP	0.97	-4.9	10.1	0.97	-5.8	13.2	0.90	-4.8	9.5
O_2 spatial	0.75	4.4	16.3	0.78	-1.4	14.5	0.73	5.6	16.3
O_2 WMP	0.91	1.7	9.8	0.83	-1.7	11.4	0.92	2.1	9.4

erage uncertainty of $5.6 \mu\text{mol kg}^{-1}$ for ΔC_{dis} (Vázquez-Rodríguez *et al.*, 2009a).

For the application case of C_{ANT} interpolation ($var = C_{ANT}$ in equation 1) the computed C_{ANT} must be normalized to a reference year (1994) using the following expression:

$$C_{ANT,j}^{1994} = \frac{C_{ANT}^{nat,1994}}{C_{ANT}^{nat,year}} C_{ANT,j}^{year} \quad (10)$$

The reason for doing this comes from the transient tracer nature of C_{ANT} and the fact that the CARINA database spans quite a long time period. Therefore, the interpolated C_{ANT} referenced to 1994 (C_{ANT}^{1994}) is computed as

$$C_{ANT}^{1994} = \frac{\sum_i C_{ANT,i}^{1994} (f_i^j)^{-1}}{\sum_i (f_i^j)^{-1}} \quad (11)$$

which is valid for both C_{ANT} reconstruction methods.

RESULTS AND DISCUSSION

To assess the quality of both types of interpolation, their results were evaluated against the WOA05 data, i.e. the interpolated *potential temperature*, *salinity*, *'NO'* and *'PO'* obtained from equation (1) were compared with the corresponding original values from WOA05 variables. Since these parameters were involved in factor's calculation of the WMP interpolation (Eqs. 2, 4), a more independent check was done by interpolating O_2 and comparing it against the WOA05

O_2 data. When the interpolation methods had been assessed, they were applied to C_{ANT} using both φC_T^o and $TrOCA$ methods, in order to get the 3D distribution and the total inventories.

The mean and standard deviation (STD) of the interpolated data minus the reference WOA05 data (S , θ , *'NO'*, *'PO'* and O_2) were computed for the whole domain and for specific sub-domains (Table 1). Also, the correlation between interpolated and reference fields was characterized with the determination coefficient (r^2). Three zones were selected depending on the general variability of water masses, namely: northern latitudes (lat $>30^\circ\text{N}$), tropical latitudes and southern latitudes (lat $<30^\circ\text{S}$). In addition, two depth levels were set with respect to the 5°C isotherms: $\theta > 5^\circ\text{C}$ and $\theta < 5^\circ\text{C}$. The 5°C isotherm represents a coarse boundary between the little-ventilated deep waters and the younger, more ventilated upper and intermediate waters. It also splits the C_{ANT} inventories in about half.

The climatological annual mean of WOA05 potential temperature along the 28°W section in WOA05 (Fig. 1) is shown in Figure 2A. The spatially-interpolated potential temperature ($\theta_{spatial}$) from the CARINA data base is close to the WOA05 data (Fig. 2B). However, some misfits do show up when the residuals are plotted (Fig. 2D). The larger residuals are located in the upper layer, where absolute values higher than 2°C are reached. In the deep water there is a better agreement, although southward of 40°S there is a large thick layer holding a systematic bias, higher than 1°C . On the other hand, the potential temperature

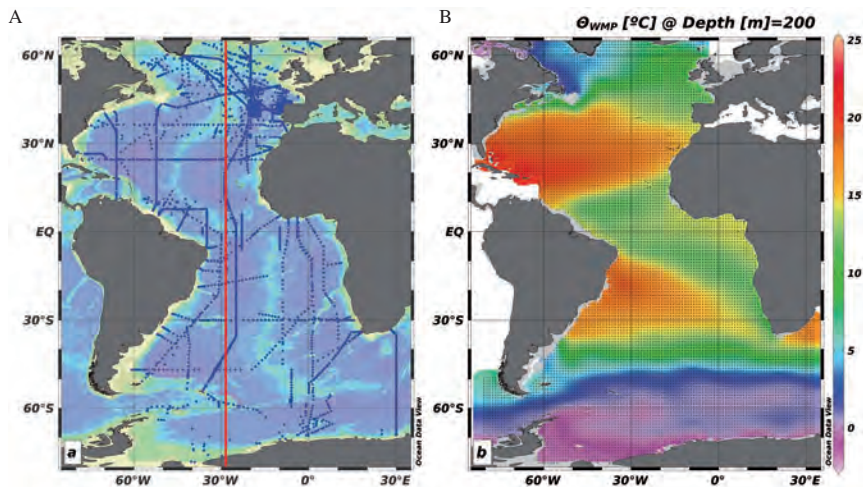


FIG. 1. – A, CARINA stations with available variables needed to estimate C_{ANT} and section 28°W plotted. B, WOA05 data grid sample with θ ($^\circ\text{C}$).

MULTIPARAMETRIC INTERPOLATION WITH WOA APPLIED TO C_{ANT} • 27

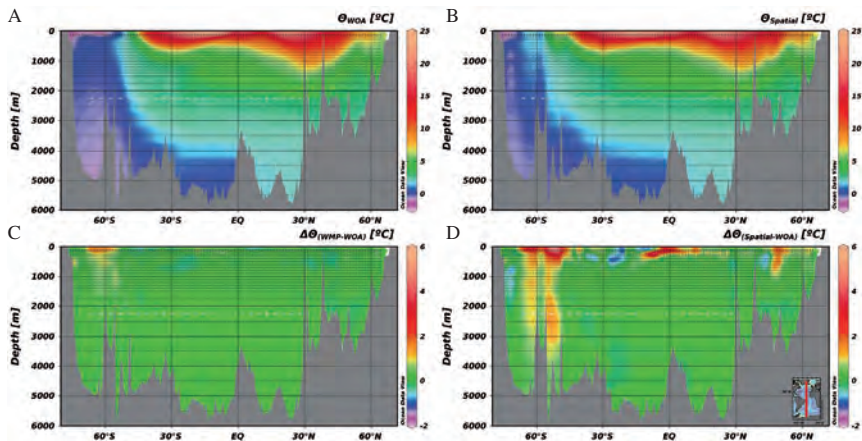


FIG. 2. – Potential temperature (°C) variability along 28°W from the WOA05 dataset (A), spatially interpolated from CARINA data (B). Residuals of WMP interpolated (C) and spatially interpolated (D) potential temperature both as interpolated minus WOA05.

from the WMP interpolation (θ_{WMP}) (Fig. 2C) shows a better agreement and lower residuals than the spatial interpolation. This time, as for $\theta_{Spatial}$, the best fit is found in deep waters, north of 45°S. Statistically speaking, the best fit (Table 1) is obtained by the WMP interpolation irrespective of whether the whole Atlantic is included or the interpolation is restricted to the sub-domains previously defined. The largest

differences between the two kinds of interpolation are located in the upper layer ($\theta > 5^\circ\text{C}$), where r^2 increases noticeably when the WMP interpolation is used. More specifically, the best fits are obtained in the tropical region and northern latitudes, whereas the southern latitudes show the worst fits. This is probably due to the lower density of data that the CARINA database has in the Southern Ocean.

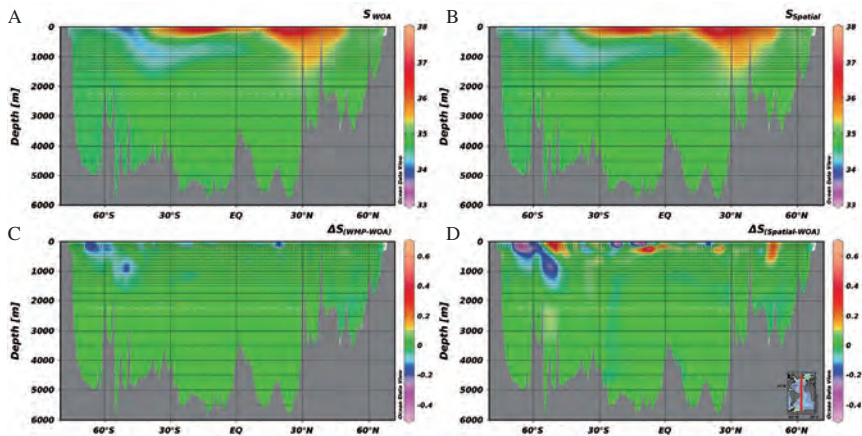


FIG. 3. – Salinity variability along 28°W from the WOA05 dataset (A), spatially interpolated from CARINA data (B). Residuals of WMP interpolated (C) and spatially interpolated (D) salinity, both as interpolated minus WOA05.

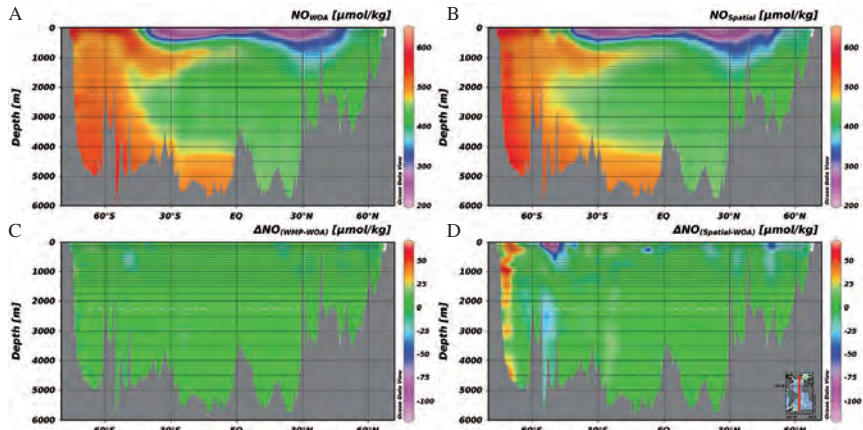


FIG. 4. – NO ($\mu\text{mol kg}^{-1}$) variability along 28°W from the WOA05 dataset (A), spatially interpolated from CARINA data (B). Residuals of WMP interpolated (C) and spatially interpolated (D) NO, both as interpolated minus WOA05.

Figure 3A shows the WOA05 climatological annual mean of salinity along the vertical section defined by the 28°W meridian (Fig. 1). As in the case of θ^{spatial} , the spatially interpolated salinity (S^{spatial}) from the CARINA database is quite close to the WOA05 reference data (Fig. 3B). Some misfits can be observed when the residuals between the spatially interpolated and WOA05 salinity values are plotted for the 28°W section (Fig. 3D). For instance, there is a considerable error

located around the salinity minimum associated with the presence of Antarctic Intermediate Water (AAIW) (Mémery *et al.*, 2000). Again, the greater residuals are found in the upper layers where absolute values higher than 0.2 psu are observed. Although in the deep waters the concordance is high in general, south of 40°S there is still the same thick layer of large biases (higher than 0.05 psu in this case) also observed with the θ^{spatial} . The WMP interpolation of salinity (S_{WMP}) (Fig. 3C) shows

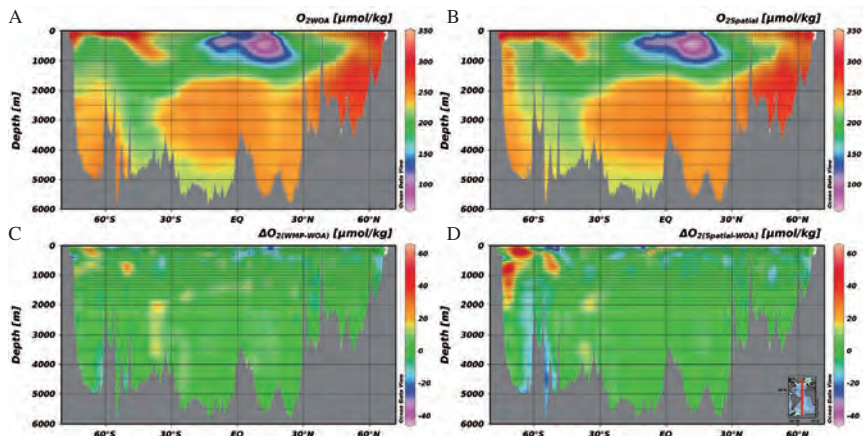


FIG. 5. – O₂ ($\mu\text{mol kg}^{-1}$) variability along 28°W from the WOA05 dataset (A), spatially interpolated from CARINA data (B). Residuals of WMP interpolated (C) and spatially interpolated (D) O₂, both as interpolated minus WOA05.

better agreement and lower residuals than the $S_{spatial}$ interpolation in all sub-domains (Table 1), proving again a superior performance of the *WMP* over the purely spatial interpolation method. Generally, both interpolation methods seem to perform better for salinity in the deep waters north of 45°S (Fig. 3D). Quantitatively speaking, the best fit to WOA05 values is obtained by the *WMP* interpolation in the tropical region (Table 1). The largest discrepancies between results from the two interpolation algorithms are found in the upper layers ($\theta > 5^\circ\text{C}$) of North Atlantic waters. Here, the r^2 obtained with the *WMP* algorithm are better by far than the ones produced by the spatial approach. However, the r^2 values obtained for S are generally slightly lower than those of θ .

The climatological annual mean of 'NO' is shown along 28°W (Fig. 4A). Since 'NO' and 'PO' are highly correlated, the 'PO' fields are not shown for the sake of conciseness. The spatial interpolation of 'NO' ($NO_{spatial}$) is close to the WOA05 data (Fig. 4B), with a most noteworthy r^2 of 0.948 (Table 1) for the whole Atlantic. The upper layers of the ocean, and particularly the Southern Ocean (SO), display the largest anomalies, which exceptionally reach offsets of $50 \mu\text{mol kg}^{-1}$ (Fig. 4D). In general, the NO_{WMP} output shows better agreement, lower standard deviations in residuals and higher determination coefficients than the $NO_{spatial}$ in every subdomain (Table 1). Unlike the θ case, the best fit is found in the upper warm layers of the Atlantic Ocean rather than in the deep ones. This is likely due to the relatively high variability of 'NO' observed in the upper layer. Conversely, the worst fit (lowest r^2), is found in the deep layers of the Southern Ocean.

The interpolated oxygen fields can be used as a test to assess the quality of the interpolation from CARINA data, given that this variable is common to both datasets and it is not used in the interpolation algorithms. The concentration of oxygen is controlled by the biological and solubility pumps. The climatological annual mean of O_2 along the 28°W section (Fig. 5A) exhibits a strong minimum in the upper layer in the tropical region caused by the remineralization of organic matter (biological pump predominance). The high values observed in the polar areas are the consequence of the high solubility of oxygen in cold surface waters (solubility pump prevails the biological processes). The spatially interpolated O_2 ($O_{2spatial}$) (Fig. 5B) closely resembles the WOA05 O_2 distribution along the 28°W vertical section (Fig. 5A). It is generally well correlated (r^2 of 0.90, Table 1) in the Atlantic Ocean, though the purely-spatial interpolation has a slight tendency to over-spline some O_2 gradients in the southern latitudes. The largest differences and lowest correlations between the observed and the $O_{2spatial}$ fields are located in the upper Atlantic layer (ranging from -25 to $20 \mu\text{mol kg}^{-1}$) and in the Southern Ocean, where offsets may reach up to $50 \mu\text{mol kg}^{-1}$ (Fig. 5D). The best fits obtained belong to the cold deep layers of the tropical region.

The O_2 *WMP* interpolation (O_{2WMP}) (Fig. 5C) is in better agreement with direct observations, has lower residuals (Fig. 5C) and has higher r^2 in general than the spatial interpolation (r^2 of 0.94 vs. r^2 of 0.90 respectively, Table 1). Unlike for the rest of domains, in deep tropical waters the *WMP* interpolation seems not to perform up to its potential, as the results from the spatial interpolation appear to be more in accordance with the observed fields. Most importantly, the *WMP* interpolation algorithm yields robust estimates in the Southern Ocean, where data coverage of the CARINA database is rather sparse.

As C_{ANT} is indistinguishable from natural CO_2 , there are no C_{ANT} benchmarks against which the estimations can be compared. Figure 6 shows the spatial and *WMP* interpolations of C_{ANT} computed using the ϕC_T° method. The general pattern of the distributions is similar to those given by Lee *et al.* (2003) and Vázquez-Rodríguez *et al.* (2009b). Sabine *et al.* (2004), using GLODAP gridded database, produced a total inventory of C_{ANT} for the Atlantic of 40 Pg-C. Previously, Lee *et al.* (2003) obtained a total inventory of 47 Pg-C by using an ungridded GLODAP database. Using CFC data, Waugh *et al.* (2006) obtained a total inventory of 48 Pg-C. A comparative study using five long transoceanic cruises was performed by Vázquez-Rodríguez *et al.* (2009b). They found that ϕC_T° and *TrOCA* turn out 55 and 51 Pg-C respectively. The results obtained here with CARINA gridded showed the same integrated inventories as Vázquez-Rodríguez *et al.* (2009b) for the ϕC_T° method, but a higher value of 58 Pg-C for the *TrOCA* method. The main differences from the GLODAP gridded data (Lee *et al.*, 2003; Key *et al.*, 2004) are located in the South Atlantic, where a large number of GLODAP estimates (obtained from the ΔC^* method, Gruber *et al.*, 1996) are negative. The GLODAP values of C_{ANT} for the Southern Ocean are lower than the ones computed here. This negative bias in the C_{ANT} estimates from the GLODAP dataset has been identified by several authors (Lo Monaco *et al.*, 2005; Waugh *et al.*, 2006; Vázquez-Rodríguez *et al.*, 2009b). The discrepancies among the interpolation methods are rather low except in the upper layers and in the Southern Ocean (Fig. 6C). The *WMP* interpolation produces higher values of C_{ANT} than the spatial method in the deep Southern Ocean ($\theta < 5^\circ\text{C}$) and lower values in the upper layers where biases reach about $\sim 8 \mu\text{mol kg}^{-1}$. In terms of C_{ANT} inventories the discrepancies are quite low and there are systematically lower (higher) values in the upper (lower) layer when the spatial interpolation is used (Table 2). The estimated total inventory of $\phi C_T^\circ C_{ANT}$ for the Atlantic is 55 Pg-C, independently of the interpolation method applied. Nevertheless, some minor discrepancies are found when the warm and cold water (θ above or below the 5°C isotherm, respectively) inventories are examined separately (Table 2).

The above-described pattern is quite similar to the one obtained when the *TrOCA* method is used to estimate C_{ANT} . The total inventory does not change too

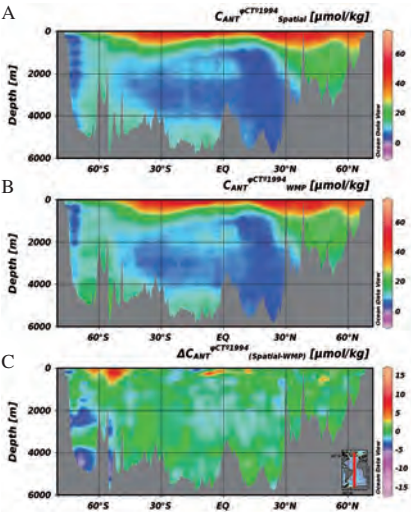


FIG. 6. – C_{ANT} ($\mu\text{mol kg}^{-1}$) determined by the φC_T^o method along 28°W using spatial interpolation (A) and WMP interpolation (B). Residuals between the two interpolations are shown (C) as spatially interpolated minus WMP interpolated.

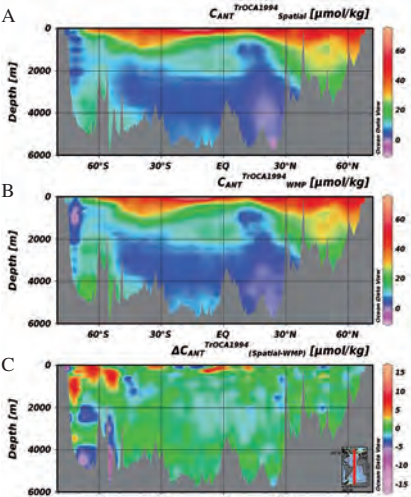


FIG. 7. – C_{ANT} ($\mu\text{mol kg}^{-1}$) determined by the TrOCA method along 28°W using spatial interpolation (A) and WMP interpolation (B). Residuals between the two interpolations are shown (C) as spatially interpolated minus WMP interpolated.

TABLE 2. – C_{ANT} inventories (Pg-C) in the Atlantic and in different latitudinal bands and layers using the TrOCA and φC_T^o methods.

C_{ANT} (Pg-C) 1994 Zone	$\theta(^{\circ}\text{C})$	φC_T^o			TrOCA		
		Spatial	WMP	diff	Spatial	WMP	diff
Lat>30°N	<5	6.9	7.0	0.1	8.1	8.1	0.0
Lat>30°N	≥5	6.0	5.8	-0.2	6.7	6.5	-0.2
Tropical	<5	12.8	13.4	0.6	12.7	13.2	0.5
Tropical	≥5	10.2	9.7	-0.5	10.5	9.9	-0.6
Lat<30°S	<5	15.5	16.3	0.8	15.9	16.7	0.8
Lat<30°S	≥5	3.7	3.1	-0.6	4.0	3.2	-0.8
Atlantic Ocean	<5	35.2	36.7	1.4	36.6	38.0	1.4
Atlantic Ocean	≥5	19.9	18.6	-1.3	21.2	19.6	-1.7
Total		55.1	55.2	0.1	57.9	57.6	-0.2

much (roughly less than 5%, Table 2) when different interpolation methods are used, although the inventories in the warm and cold layers for the whole Atlantic Ocean are around 1.5 Pg-C different depending on the interpolation method used. Using the TrOCA method produces a slightly higher C_{ANT} inventory. The TrOCA method gives higher values in the surface layer (higher penetration) and in the deep Northern North Atlantic (Fig. 7). In comparison, the φC_T^o method gives slightly higher values in practically all deep water masses and in the Southern Ocean, except for the Antarctic Bottom Water (AABW). Again, the major differences found in the southern latitudes are a consequence of the low density of carbon system data in this region.

Future work will be needed to improve the interpolation method and obtain an uncertainty assessment, and to iterate back and forth to the original data following a Barnes schema to fine-tune the interpolation. After these improvements have been achieved and with uncertainties available, the enhanced interpolation method could be used to interpolate the parameters available in CARINA and GLODAP, in order to provide an enhanced gridded product. Also, more C_{ANT} estimation techniques such as TTD and ΔC^* can be incorporated and applied to the CARINA database so that their inventories will be obtained.

CONCLUSIONS

The WMP interpolation method offers improvements compared with a traditional spatial gridding, and even with an objective analysis spatial gridding. By using an auxiliary database (WOA05) constructed with more resolution data on bio-geochemical conservative parameters, the WMP method has more information for the gridding task than any other exclusively spatial alternative.

The total inventory of C_{ANT} (referred to 1994) for the Atlantic Ocean is estimated to be about 55-58 Pg-C depending on the C_{ANT} estimation technique applied (φC_T^o or TrOCA, respectively). The interpolation methods used here (spatial and WMP) do not have significant effects on the estimates of C_{ANT} total inventories, due to compensation effects between domains. Nevertheless, there exist some minor differences in the

results obtained by the different interpolation methods. The WMP interpolation method performs better than the spatial one, particularly in regions with less density of initial data (most importantly the Southern Ocean) from the CARINA dataset. Finally, the differences between the interpolation methods transcend to the realm of C_{ANT} estimation above and below the 5°C isopleth. The spatial method tends to produce lower (higher) C_{ANT} values in the water below (above) the isotherm of 5°C than the WMP interpolation method.

ACKNOWLEDGEMENTS

We would like to extend our gratitude to the chief scientists, scientists and crew who participated and offered their dedication to the oceanographic cruises used in this study. This work was developed and funded by the European Commission within the 6th Framework Programme (EU FP6 CARBOOCEAN Integrated Project, Contract no. 511176), MEC (CTM2006-27116-E/MAR) and by the Xunta de Galicia within the INCITE framework (M4AO project PGIDIT07PKB402153PR).

We also wish to thank the editor and the two anonymous reviewers, whose comments greatly contributed to improving and focusing the manuscript.

REFERENCES

- Anderson, L.A. and J.L. Sarmiento. – 1994. Redfield Ratios of Remineralization Determined by Nutrient Data Analysis. *Global Biogeochem. Cycles*, 8: 65-80.
- Antonov, J.I., R.A. Locarnini, T.P. Boyer, A.V. Mishonov, H.E. Garcia and S. Levitus. – 2006. World Ocean Atlas 2005, Volume 2: Salinity. S. Levitus. NOAA Atlas NESDIS 62, U.S. Government Printing Office, Washington, D.C.: 182.
- Bennett, A.F. – 1992. *Inverse methods in physical oceanography*. Cambridge University Press.
- Bretherton, F.P., R.E. Davis and C.B. Fandry. – 1976. A technique for objective analysis and design of oceanographic experiments applied to MODE-73. *Deep-Sea Res.*, 23: 559-582.
- Broecker, W.S. – 1974. "NO", a conservative water-mass tracer. *Earth Planet. Sci. Lett.*, 23.
- Canadell, J.G., C. Le Quéré, M.R. Raupach, C.B. Field, E.T. Buitenhuis, P. Ciais, T.J. Conway, N.P. Gillett, R.A. Houghton and G. Marland. – 2007. Contributions to accelerating atmospheric CO₂ growth from economic activity, carbon intensity, and efficiency of natural sinks. *Proc. Natl. Acad. Sci.*, 104: 18866.
- Emery, W.J. and R.E. Thomson. – 2001. *Data analysis methods in physical oceanography*. Elsevier Science.
- Freeland, H.J. and W.J. Gould. – 1976. Objective analysis of meso-scale ocean circulation features. *Deep-Sea Res.*, 23: 915-923.
- Gandin, L.S. and R. Hardin. – 1965. *Objective analysis of meteorological fields*. Israel program for scientific translations Jerusalem.
- Garcia, H.E., R.A. Locarnini, T.P. Boyer, J.I. Antonov and S. Levitus. – 2006a. World Ocean Atlas 2005, Volume 3: Dissolved Oxygen, Apparent Oxygen Utilization, and Oxygen Saturation. S. Levitus. NOAA Atlas NESDIS 63, U.S. Government Printing Office, Washington, D.C.: 342.
- Garcia, H.E., R.A. Locarnini, T.P. Boyer, J.I. Antonov and S. Levitus. – 2006b. World Ocean Atlas 2005, Volume 4: Nutrients (phosphate, nitrate, silicate). S. Levitus. NOAA Atlas NESDIS 64, U.S. Government Printing Office, Washington, D.C.: 396.
- Gruber, N., J.L. Sarmiento and T.F. Stocker. – 1996. An improved method for detecting anthropogenic CO₂ in the oceans. *Global Biogeochem. Cycles*, 10: 809-837.
- Hedges, J.I., W.A. Clark and G.L. Cowie. – 1988. Organic matter sources to the water column and superficial sediments of a marine bay. *Limnol. Oceanogr.*, 33: 1116-1136.
- Heinze, C. – 2004. Simulating oceanic CaCO₃ export production in the greenhouse. *Geophys. Res. Lett.*, 31: L16308.
- Hoppema, M., A. Velo, S. van Heuven, T. Tanhua, R.M. Key, X. Lin, D.C.E. Bakker, F.F. Pérez, A.F. Ríos, C. Lo Monaco, C.L. Sabine, M. Alvarez and R.G.J. Bellerby. – 2009. Consistency of cruise data of the CARINA database in the Atlantic sector of the Southern Ocean. *Earth Syst. Sci. Data*, 1: 63-75.
- Jallicee, J.B. and D.R. Hamilton. – 1977. Objective analysis and classification of oceanographic data. *Tellus*, 29.
- Joos, F., G.K. Plattner, T.F. Stocker, O. Marchal and A. Schmittner. – 1999. Global warming and marine carbon cycle feedbacks on future atmospheric CO₂. *Science*, 284: 464.
- Key, R.M., A. Kozyr, C.L. Sabine, K. Lee, R. Wanninkhof, J.L. Bullister, R.A. Feely, F.J. Millero, C. Mordy and T.H. Peng. – 2004. A global ocean carbon climatology: Results from Global Data Analysis Project (GLODAP). *Global Biogeochem. Cycles*, 18, GB4031.
- Key, R.M., T. Tanhua, A. Olsen, M. Hoppema, S. Jutterström, C. Schirnick, S. van Heuven, A. Kozyr, X. Lin, A. Velo, D.W.R. Wallace and L. Mintrop. – 2010. The CARINA data synthesis project: Introduction and overview. *Earth Syst. Sci. Data*, 2: 105-121.
- Kortzinger, A., J.I. Hedges and P.D. Quay. – 2001. Redfield ratios revisited: Removing the biasing effect of anthropogenic CO₂. *Limnol. Oceanogr.*, 964-970.
- Le Quéré, C., C. Rodenbeck, E.T. Buitenhuis, T.J. Conway, R. Langenfelds, A. Gomez, C. Labuschagne, M. Ramonet, T. Nakazawa and N. Metzl. – 2007. Saturation of the Southern Ocean CO₂ sink due to recent climate change. *Science*, 316: 1735.
- Lee, K., S.D. Choi, G.H. Park, R. Wanninkhof, T.H. Peng, R.M. Key, C.L. Sabine, R.A. Feely, J.L. Bullister and F.J. Millero. – 2003. An updated anthropogenic CO₂ inventory in the Atlantic Ocean. *Global Biogeochem. Cycles*, 17: 1116.
- Lo Monaco, C., C. Goyet, N. Metzl, A. Poisson and F. Touratier. – 2005. Distribution and inventory of anthropogenic CO₂ in the Southern Ocean: Comparison of three data-based methods. *J. Geophys. Res.*, 110: 9.
- Locarnini, R.A., A.V. Mishonov, J.I. Antonov, T.P. Boyer, H.E. Garcia and S. Levitus. – 2006. World Ocean Atlas 2005, Volume 1: Temperature. S. Levitus. NOAA Atlas NESDIS 61, U.S. Government Printing Office, Washington, D.C.: 182.
- Matear, R.J., C.S. Wong and L. Xie. – 2003. Can CFCs be used to determine anthropogenic CO₂? *Global Biogeochem. Cycles*, 17: 1013.
- Mémery, L., M. Arhan, X.A. Alvarez-Salgado, M.J. Messias, H. Mercier, C.G. Castro and A.F. Ríos. – 2000. The water masses along the western boundary of the south and equatorial Atlantic. *Progr. Oceanogr.*, 47: 69-98.
- Pérez, F.F., C. Mourino, F. Fraga and A.F. Ríos. – 1993. Displacement of water masses and remineralization rates off the Iberian Peninsula by nutrient anomalies. *J. Mar. Res.*, 51: 869-892.
- Ríos, A.F., F. Fraga and F.F. Pérez. – 1989. Estimation of coefficients for the calculation of "NO", "PO" and "CO", starting from the elemental composition of natural phytoplankton. *Sci. Mar.*, 53: 779-784.
- Sabine, C.L., R.M. Key, K.M. Johnson, F.J. Millero, A. Poisson, J.L. Sarmiento, D.W.R. Wallace and C.D. Winn. – 1999. Anthropogenic CO₂ inventory of the Indian Ocean. *Global Biogeochem. Cycles*, 13: 179-198.
- Sabine, C.L., R.A. Feely, N. Gruber, R.M. Key, K. Lee, J.L. Bullister, R. Wanninkhof, C.S. Wong, D.W.R. Wallace and B. Tilbrook. – 2004. The oceanic sink for anthropogenic CO₂. *Science*, 305: 367-371.
- Tanhua, T., A. Olsen, M. Hoppema, S. Jutterström, C. Schirnick, S. Van Heuven, A. Velo, X. Lin, A. Kozyr, M. Alvarez, D.C.E. Bakker, P. Brown, E. Falck, E. Jeansson, C. Lo Monaco, J. Olafsson, F.F. Perez, D. Pierrot, A.F. Rios, C.L. Sabine, U. Schuster, R. Steinfeldt, I. Stendardo, L.G. Anderson, N.R. Bates, R.G.J. Bellerby, J. Blindheim, J.L. Bullister, N. Gruber, M. Ishii, T. Johannessen, E.P. Jones, J. Köhler, A. Kortzinger, N. Metzl, A. Murata, S. Musielewicz, A.M. Omar, K.A. Ölsön, M. de la Paz, B. Pfeil, F. Rey, M. Rhein, I. Skjelvan, B. Tilbrook, R. Wanninkhof, L. Mintrop, D.W.R. Wallace and R.M. Key. – 2009. CARINA Data Synthesis Project. ORNL/CDIAC-157, NDP-091, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee, 37831-6335.

- Tanhua, T., R. Steinfeldt, R.M. Key, P. Brown, N. Gruber, R. Wanninkhof, F.F. Pérez, A. Körtzinger, A. Velo, U. Schuster, S. van Heuven, J.L. Bullister, I. Stendardo, M. Hoppema, A. Olsen, A. Kozyr, D. Pierrot, C. Schimick and D.W.R. Wallace. – 2010. Atlantic Ocean CARINA data: overview and salinity adjustments. *Earth Syst. Sci. Data*, 2: 17-34.
- Thomas, H. and V. Ittekkot. – 2001. Determination of anthropogenic CO₂ in the North Atlantic Ocean using water mass ages and CO₂ equilibrium chemistry. *J. Mar. Syst.*, 27: 325-336.
- Touratier, F., L. Azouzi and C. Goyet. – 2007. CFC-11, $\Delta^{14}\text{C}$ and ^3H tracers as a means to assess anthropogenic CO₂ concentrations in the ocean. *Tellus B*, 59: 318-325.
- U.S. Department of Commerce, N.O.a.A.A., National Geophysical Data Center. – 2006. *2-minute Gridded Global Relief Data (ETOPO2v2)*.
- Vázquez-Rodríguez, M., X.A. Padin, A.F. Ríos, R.G.J. Bellerby and F.F. Pérez. – 2009a. An upgraded carbon-based method to estimate the anthropogenic fraction of dissolved CO₂ in the Atlantic Ocean. *Biogeosci. Discuss.*, 6: 4527-4571.
- Vázquez-Rodríguez, M., F. Touratier, C. Lo Monaco, D.W. Waugh, X.A. Padin, R.G.J. Bellerby, C. Goyet, N. Metzl, A.F. Ríos and F.F. Pérez. – 2009b. Anthropogenic carbon distributions in the Atlantic Ocean: data-based estimates from the Arctic to the Antarctic. *Biogeosciences*, 6: 439-451.
- Waugh, D.W., T.M. Hall, B.I. McNeil, R. Key and R.J. Matear. – 2006. Anthropogenic CO₂ in the oceans estimated using transit time distributions. *Tellus B*, 58: 376-389.

Received November 1, 2008. Accepted July 26, 2010.

Published online November 13, 2010.

Capítulo III

Total alkalinity estimation using MLR and neural network techniques



Contents lists available at SciVerse ScienceDirect

Journal of Marine Systems

journal homepage: www.elsevier.com/locate/jmarsys


Total alkalinity estimation using MLR and neural network techniques

A. Velo ^{a,*}, F.F. Pérez ^a, T. Tanhua ^b, M. Gilcoto ^a, A.F. Ríos ^a, R.M. Key ^c
^a Instituto de Investigaciones Mariñas (CSIC), Eduardo Cabello 6, 36208 Vigo, Spain

^b GEOMAR Helmholtz Centre for Ocean research Kiel, Marine Biogeochemie, Kiel, Germany

^c Princeton University, Program in Atmospheric and Oceanic Science, Forrestal Campus/Sayre Hall, Princeton, NJ 08544, USA

ARTICLE INFO

Article history:

Received 25 August 2011

Received in revised form 27 August 2012

Accepted 5 September 2012

Available online 12 September 2012

Keywords:

Seawater alkalinity

Ocean carbonate system

MLR techniques

Neural network techniques

Data quality control

ABSTRACT

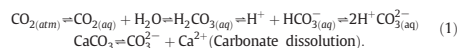
During the last decade, two important collections of carbon relevant hydrochemical data have become available: GLODAP and CARINA. These collections comprise a synthesis of bottle data for all ocean depths from many cruises collected over several decades. For a majority of the cruises at least two carbon parameters were measured. However, for a large number of stations, samples or even cruises, the carbonate system is under-determined (i.e., only one or no carbonate parameter was measured) resulting in data gaps for the carbonate system in these collections. A method for filling these gaps would be very useful, as it would help with estimations of the anthropogenic carbon (C_{ant}) content or quantification of oceanic acidification. The aim of this work is to apply and describe, a 3D moving window multilinear regression algorithm (MLR) to fill gaps in total alkalinity (A_T) of the CARINA and GLODAP data collections for the Atlantic. In addition to filling data gaps, the estimated A_T values derived from the MLR are useful in quality control of the measurements of the carbonate system, as they can aid in the identification of outliers. For comparison, a neural network algorithm able to perform non-linear predictions was also designed. The goal here was to design an alternative approach to accomplish the same task of filling A_T gaps. Both methods return internally consistent results, thereby giving confidence in our approach.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

GLODAP (Key et al., 2004; Sabine et al., 2005) and CARINA (Key et al., 2010; Tanhua et al., 2009) are data collections of ocean bottle samples from the synthesis of many oceanic cruises. GLODAP provides reasonable global coverage except for the Arctic. CARINA focuses on the Atlantic with particular emphasis on the North Atlantic and Arctic Ocean. Both collections are the result of international efforts to summarise all available data into compiled products.

Total dissolved inorganic carbon (C_T), total alkalinity (A_T), total hydrogen ion concentration (pH) and fugacity of CO_2 are the four measured parameters of the carbonate system in seawater. A large fraction of the samples in these databases has at least one carbon parameter measured and their merging results in the distribution of measured parameters, as shown in Fig. 1. The seawater carbonate system is governed by the thermodynamics of the carbonate system through the following equation (Eq. (1)) (Dickson, 2010; Millero, 1995; Zeebe and Wolf-Gladrow, 2001):



The system is defined by having any two of the four possible parameters (Lewis et al., 1998; Millero, 1995; van Heuven et al., 2009).

Among the carbon system parameters, A_T can be estimated most accurately due to its low variability in the ocean (Wolf-Gladrow et al., 2007; Zeebe and Wolf-Gladrow, 2001). A_T is supposed to be only marginally affected by increasing C_T concentration, at least at the current decadal scale, because the carbonate system tends to buffer the alkalinity when dissolving CO_2 in contemporary seawater (Ilyina et al., 2009; Riebesell et al., 2010). Most of the A_T variability is known and is well correlated with salinity, silicate and even temperature (Lee et al., 2006; Millero et al., 1998; Wallace, 1995). This is advantageous for this work, as the full dataset of available A_T data from old cruises performed in the 1970s to the most recent can be used directly as a climatology for the multilinear regression (MLR) computations. There is no need for partitioning the dataset in a time series or for introducing the time parameter in estimations, which would result in fewer available data, or less precision for estimations. However, alkalinity measurements performed without using certified reference materials (CRMs) have proven to be less accurate, which is true for most data obtained prior to the mid-1990s.

GLODAP bottle cruise data for the Atlantic region were used together with CARINA bottle cruise data for the Atlantic and Southern Ocean regions. Both data collections (i.e., the data products with adjustments applied to certain parameters, see Sabine et al. (2005) for the GLODAP data and Tanhua et al. (2010) for the CARINA data) were merged and outlier filtering was performed to prepare the dataset. The existence

* Corresponding author. Tel.: +34 986231930.
E-mail address: avelo@iim.csic.es (A. Velo).

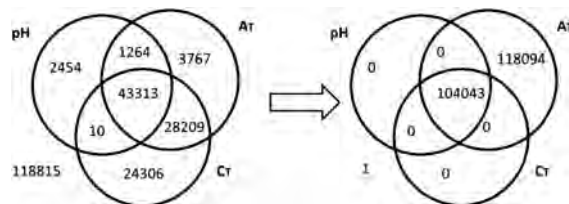


Fig. 1. Venn diagram with carbon data availability before and after MLR calculations; the numbers within the different fields are the number of samples that have been assigned with a value for each of the parameters pH, AT, CT, a combination of them, or no carbonate system data (lower left). Left panel: original data from the CARINA/GLODAP Atlantic merge; right panel: after adding the additional data obtained in this study.

of depth, temperature and salinity data for each sample was required and so, samples without these variables were removed. The absence or bad data of nitrate and phosphate samples were recovered by using the same multilinear regression presented here for alkalinity, once the algorithm was fine-tuned. Multiparametric spatial/WMP interpolation techniques (Velo et al., 2010) with the World Ocean Atlas 2009 (WOA09) (García et al., 2009) were used to recover silicate samples where no data were available. The Results and discussion section (3.1) provides details on these procedures.

Two very different techniques are being proposed here to infer the A_T values. Firstly, a customised 3D moving window MLR method (hereinafter named 3DwMLR), fine-tuned and focused for this purpose; and secondly, a customised neural network, also parameterised and fine-tuned for this task. Both alternatives provide fundamentally different approaches to the same problem and thus, a comparison of their results was used as a validation tool for the analysis.

2. Methodology

2.1. The 3D moving window MLR (3DwMLR) estimation

The classical multilinear regression (MLR, or LM) scheme allows modelling of the relationship between a response variable and some predictors by using a linear equation combining these predictors. The problem with using a classical MLR applied to the full CARINA and GLODAP dataset is that not all A_T natural variability can be modelled this way and large errors may result. The A_T relationship with salinity and temperature could be assumed to be linear, but only when data are restricted to specific confined areas in basins and mainly for surface waters (Lee et al., 2006; Millero et al., 1998).

The objective of this work is to enhance the results of the classical MLR by combining the following techniques:

1. Use a 3D moving window to define a data box around the node for which we are trying to calculate A_T and to perform the regression using only the data inside that box.
2. Use a robust regression algorithm instead of ordinary least squares regression.
3. Include other parameters in the regression in addition to salinity and temperature (Lee et al., 2006; Millero et al., 1998), which have influence, directly or indirectly, on the natural variability of A_T ; these can be silicate, nitrate, phosphate and oxygen.

2.1.1. 3D moving window

An algorithm was designed to extract a pool of data from a box around each bottle sample. This box is referred to as the 3D moving window as its central node changes for and is centred on each sample. All samples of the merged GLODAP–CARINA dataset were used and all of them have their specific data window in order to make the MLR calculations. The width and depth of the data window were chosen according to the previous procedures of Velo et al. (2010)

and multiple comparisons of different parameterisations were done to improve the results.

Summarising them; variable-width and fixed-width moving windows have been tried, both with the requirement of a minimum number of samples inside the window to perform the MLR. The variable-width tests started with a radius of 2 latitudinal degrees around the node, allowing the algorithm to increase the size degree by degree until reaching 100 samples around the node. For depth, a direct depth-dependent equation of $\text{height} = 150 + 0.15 \cdot [\text{sample depth}]$ around the node was used, allowing the algorithm to grow 5% in depth for each degree of latitudinal increase. Fixed-width tests used a wider area of 5, 10 or even 20 latitudinal degrees around the node and the same equation for depth.

Variable-width parameterisation more closely mimics the nearby sample values, but is affected more by the scarcity of data. However, fixed-width settings tend to smooth the natural variability or even add noise if the window is too wide.

A mostly-fixed moving window of 10° in latitude has proven to be the most consistent parameterisation for A_T calculations with the CARINA–GLODAP merge for the Atlantic Ocean, compared with observations. The term mostly-fixed means that the algorithm was allowed to increment the size of its moving window only when less than 100 neighbouring samples were found in the box, which only happened for 5324 of a total 222,136 estimated A_T values (2.4% of samples). The increments were made in steps of 1° in latitude and 5% in depth, up to a maximum window of 20° ($+63\%$ in depth). By using this combination of settings, the A_T dataset gets the most consistent result among the presented options, allowing for the recovery of large amounts of data.

2.1.2. Robust regression

A robust regression algorithm (Filzmoser et al., 2007; Hubert, 2008) was used as the MLR technique. The robust regression combines an ordinary regression algorithm with outlier detection by performing iterations to re-weight predictors. This technique is also known as interactive reweighting least squares (IRLS) (Green, 1984). Re-weighting is performed using a bi-square weighting function over residuals. Thus, the inferred A_T is more consistent compared with observations than when an ordinary least-squares regression is used.

In ordinary least-squares (OLS) regression it is a critically important task to keep input values under strict control, as the OLS regression is very sensible to outliers. By using a robust regression technique, the weighting function is able to detect most of these outliers by reiterating and thus, discarding them or giving them a minimum impact on the regression. A bi-square function with a cut-off value of 4.685/residuals (Holland and Welsch, 1977) was used for weighting in each iteration step. Weights decay quickly as residuals deviate from zero and the points are discarded (weight 0) when residuals exceed the cut-off value. The robust regression technique remains statistically robust with up to 50% of outliers present

in the dataset, whereas for the OLS regression, the presence of one single outlier influences the result.

2.1.3. Selection of predictor parameters

As predictor parameters pressure, potential temperature (θ), salinity (S), nitrate, phosphate, silicate and oxygen were used. The main objective of the addition of these parameters was to incorporate as much of the natural variability as possible in order to get better predictions for each regional A_T regression. These parameters have been chosen as they are related to alkalinity and their abundance is larger than the carbon parameters in the dataset. The relationship between A_T and both S and θ is well-known (Lee et al., 2006; Millero et al., 1998). The relationship of A_T with nutrients (nitrate, phosphate, silicate) is related to organic matter synthesis and re-mineralisation that adds or removes these ions from dissolution and thus, affects A_T (Chen and Millero, 1979; Chen and Pytkowicz, 1979). Furthermore, it can be considered that the sum of the A_T , phosphate and nitrogen species remains constant during the re-mineralisation of organic matter (Brewer et al., 1975; Fraga and Álvarez-Salgado, 2005). Oxygen also influences A_T through its use in organic synthesis and re-mineralisation processes. The pressure relationship with A_T can also be related to carbonate formation and its dissolution at depth in the ocean. All these processes have been treated extensively in the literature (Dickson, 2010; Millero, 2007; Zeebe and Wolf-Gladrow, 2001).

The influence of these parameters on the A_T regression has also been analysed from a statistical point of view, to verify that the parameter selection is adequate and that all parameters provide information to the model. It should be noted that the ultimate purpose of this work is the prediction of A_T values with the most realistic fit, with less emphasis on the search for the relative significance of each predictor in A_T . The existence of correlations and collinearity among a number of predictors is well known by the authors, but from a prediction point of view, they should not be removed if their presence improves the prediction. Moreover, we are using a 3D moving window and the collinearities may change from window to window. In addition, the relative importance of the parameters may vary for each window fit resembling the oceanic variability. Therefore, we do not discard a priori any parameter; the MLR of each window will adapt the relative contribution for each parameter to the local conditions of the 3D window.

To assess the overall collinearity among predictors, we have performed two statistical analyses. First, variance inflation factors (VIF) were calculated for a multilinear regression with the full dataset, following the procedures of Juranek et al. (2009) and Kim et al. (2010). The results are presented in Table 1, where each row shows the statistics for each MLR model considering as predictors the parameter of that row and below. The VIF factor indicates collinearity for values greater than 5–10 (Kutner et al., 2004) and very high collinearity is found when nitrate and phosphate are used. It is well-known that statistically, nitrate correlates highly with phosphate (Redfield, 1934), but it is also known that the N:P relationship has biogeochemical

sources of variability (Anderson and Sarmiento, 1994; Geider and La Roche, 2002). Thus, as both presented models make use of geographical information, the inclusion of both phosphate and nitrate can improve the predictions.

Second, Akaike Information Criterion (AIC) values were also calculated for a full model with all the indicated predictors and for MLR models excluding the highest VIF values. In Table 1, the AIC is presented as ΔAIC , i.e., the AIC for the MLR model of that row minus the AIC for the MLR with all (7) predictors (first table row). Lower ΔAIC values indicate better significance (Anderson and Burnham, 2002) of the model. R squared values have also been calculated for each subset combination of predictors (Miller, 1984), indicating the best fit when all seven parameters are used.

Inside each 3D moving window, the distribution of predictor values around the node value was tested for each parameter. If this node value was statistically inside the data cloud, only samples with more than three times the standard deviation (STD) from the node value were discarded. Otherwise, if most of the cloud of samples around this node was deviated, the entire predictor was discarded for the MLR in that node, in order to avoid extrapolation biases in the predicted A_T . The algorithm was designed with freedom to remove up to two parameters, giving no predicted A_T if more than two parameters had to be discarded.

The relationship between the estimated A_T and the predictors has been assumed linear within each 3D window, i.e., the ranges of variation of A_T and predictors in the windows are small enough to allow linear approximations. However, the method to select predictors inside the 3D windows assures that the predictor node values, those used to calculate the interpolated A_T , are inside the ranges of the predictors in the 3D window.

2.2. Neural network estimation

For comparison with the previous algorithm, a neural network was constructed and tested to predict A_T in the same area and with the same predictors, plus the geographical information. Two main steps were needed: 1) design, configure and implement an adequate network model and 2) tune the model parameters to optimise performance and results, i.e., what commonly is referred to as training. An overview of these procedures can be found in Sarle (1997).

2.2.1. Design

A feed-forward multilayer network with a hidden log-sigmoidal layer and a linear output was designed and configured (Gardner and Dorling, 1998). The advantage of such architecture is that the log-sigmoidal layer can learn the non-linear relationships between predictors and results, whereas the final linear output layer deals with linear variability (Bishop, 1995; Hagan et al., 1996). Each node in a neural network applies a transfer function from its inputs to its outputs by weights tuned in the training step. A schematic diagram is shown in Fig. 2. The combination and superposition of results from all simple nodes and layers is the key for modelling very complex non-linear problems (Gardner and Dorling, 1998).

Latitude, longitude (in -180° to 180° scale) and depth were used as predictor parameters to inform the network about the sample position, while θ , S , phosphate, nitrate, silicate and oxygen informed about the seawater properties. Apart from sample position, this choice matches the parameters chosen for the 3DwMLR estimation.

There is no direct way to establish the optimal number of nodes in the hidden layer (Sarle, 1997); the typical evaluation procedure is to try different network sizes and evaluate their confidence with observations. Network models with a range from 2^1 to 2^9 hidden neurons (units) were tested. A network size of 64 hidden units proved to be the best compromise between underestimation and overfitting and thus, was deemed to be the optimal network size.

Table 1
Statistical test on collinearity and relevance of predictors.

Parameter	VIF(7)	VIF(6)	VIF(5)	VIF(4)	No. par. used	ΔAIC	R^2
Nitrate	117.15				7	0.0	0.930
Phosphate	116.26	17.18			6	1539.1	0.929
Theta	14.41	10.88	5.54		5	9883.7	0.923
Oxygen	5.76	4.97	1.29	1.19	4	13249.6	0.921
Salinity	5.31	4.97	3.57	1.46	3	15282.8	0.919
Silicate	3.67	3.50	1.77	1.59	2	273075.5	0.039
Pressure	1.93	1.86	1.75	1.32	1	276923.6	0.002

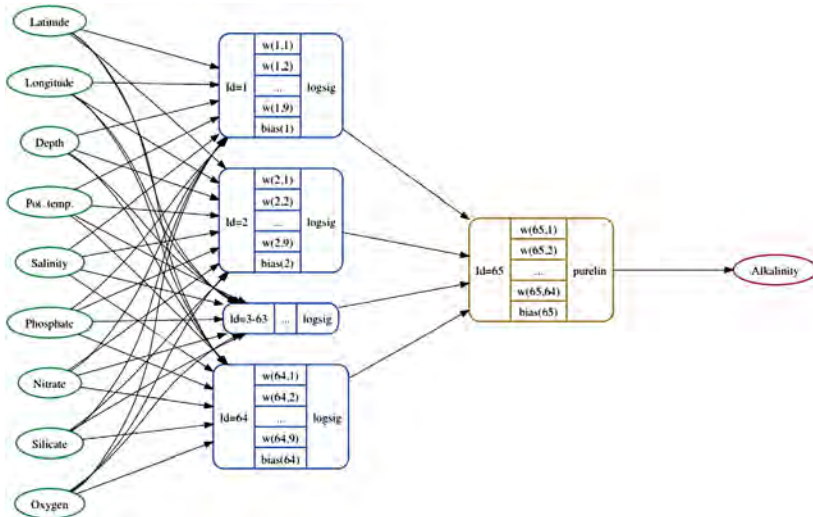


Fig. 2. Schematic diagram of the feed-forward neural network designed for A_T prediction. Nodes with $Id = 1-64$ represent the first log-sigmoidal hidden layer (logsig) and node with $Id = 65$ represents the pure linear layer (purelin). Nodes 3–63 are represented by dots in the middle box. Values w_{1-9} and w_{1-64} together with their bias are the weights established thorough training and that makes the network useful to predict.

The results for a number of different training techniques are presented in Fig. 3.

2.2.2. Tuning

Artificial feed forward neural networks need training and they are commonly trained by back-propagation techniques (Hagan et al., 1996; Sarle, 1997). These techniques feed the network iteratively and in batch-mode with values of the training dataset and back-propagate the errors (deltas) of the results through the weights determined in the network. There are many training techniques in the literature, which differ on how the weights are modified by the back-propagation. Some of these training techniques were analysed and tested to tune the design of the network. Fig. 3 shows the STD

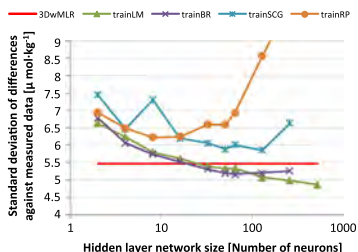


Fig. 3. Comparison of some different neural network training techniques and MLR with the measured data.

of the differences between the predicted A_T and the observations, plotted for the proposed different training techniques and against increasing network sizes (number of neurons in the hidden layer). As can be seen, the performance of techniques such as Resilient Backpropagation (trainRP), Scaled Conjugate Gradient (trainSCG) or Bayesian Regularisation (trainBR) quickly degrades as the size of the hidden layer exceeds a number of neurons (Hagan et al., 1996). The best training technique, according to observations, has proven to be the Levenberg–Marquardt (trainLM) method (Hagan and Menhaj, 1994), using a network size of 64 neurons in the hidden layer. Larger network sizes have revealed a tendency to overfit (Bishop, 1995; Hagan et al., 1996).

One of the main problems with neural networks is overfit. Neural networks are very powerful and by increasing the number of neurons in a non-linear layer, they can eventually model perfectly (producing null residuals). Therefore, care must be taken not to overfit the network or the model will also incorporate the measurement errors. This can be detected by testing or validating the network with data not used in the learning steps of the training process.

To train a neural network, some data are needed to validate the improvement in each learning iteration and other data for testing the results. Thus, the dataset has to be partitioned. The most consistent scheme consisted of using an arbitrary selection of data for the partitions and re-doing the arbitrary partitioning in each iteration. 80% of the dataset was used for learning procedures, 15% for validation and 5% for testing. However, by design, the Bayesian regularisation training technique always uses the whole dataset for learning.

Mean squared error was chosen as the convergence statistic in each iteration for all training methods, except Bayesian regularisation, that used the sum of squared errors. These analyses are designed to maximise the fit between learning and validation partitions by iteratively tuning the weights.

As the convergence approaches a minimum, the learning process stops and the neural network saves the weighting data information for use in predicting A_T from inputs (Fig. 2).

3. Results and discussion

3.1. Initial quality control of input data

As a starting point, the GLODAP and CARINA databases were merged, providing a total of 282,853 samples available for the Atlantic basin. From these, only 84,734 initially had A_T values (29.96%).

An initial check resulted in 18,063 samples being discarded as they lacked pressure or depth, θ or S data. Additionally, 32,910 samples without oxygen data were removed.

Measurements of nutrient parameters were missing for a large number of samples. In order to recover some of these data points we used the World Ocean Atlas (WOA09) dataset as follows:

Silicate: From 14,085 samples without silicate, 4661 new values were incorporated using a combined Spatial/WMP interpolation (Velo et al., 2010) applied to the WOA09 gridded climatology. The climatology of silicate was interpolated to cruise sample nodes and θ , S , latitude, longitude and depth were used for the parametric distances in the interpolation algorithm. The remaining 9424 samples without silicate were removed from the dataset.

Nitrate and phosphate: 3DwMLR technique was used to obtain nitrate or phosphate in samples where their counterpart phosphate or nitrate was available. For these MLR, the respective available phosphate or nitrate parameter was used as the predictor parameter and also oxygen, θ , S and distances were incorporated as input in the MLR. For samples without these nutrients, a 3DwMLR with only oxygen, θ , S and distances was used. As a result, 6780 samples without nitrate and 15,220 samples without phosphate were filled and no data without nitrate or phosphate were left in the database.

Some further quality control was performed for the nitrate and phosphate dataset of measurements. The 3DwMLRs indicated above, were also applied to samples with measured data in order to compare predicted with measured data for these parameters. The difference between the measured data with the two MLR was taken as quality criterion for keeping or replacing the measured value. This check yielded 6268 nitrate samples and 4146 phosphate samples replaced by 3DwMLR values.

As a final quality control of input data, station profiles were manually supervised to identify outliers and were removed from the dataset for the A_T calculations. Because of this control, 4425 sample values were removed.

3.2. Alkalinity estimation

After the initial checks and quality control, the dataset was reduced to 222,138 samples useful for calculations, with 76,553 samples having observations of A_T . Then, both the 3DwMLR and the neural network algorithm were used as stated in the Methodology section to predict A_T for the full dataset. Both algorithms were able to generate 222,136 estimated values of A_T , i.e., for all but two single samples.

3.3. Quality control for observed A_T data

The estimated A_T data was used to perform a quality control of the observations, comparing them with the values predicted by the algorithm. The STD of the difference between A_T observations and the predicted A_T by both methods was used as quality control criterion for the observations. The magnitude of this STD is a very good

reference for the overall performance of the algorithms to predict the A_T data. An STD value of $5.4 \mu\text{mol kg}^{-1}$ was obtained for the 3DwMLR method and a value of $5.2 \mu\text{mol kg}^{-1}$ for the chosen neural network trained with the Levenberg–Marquardt method and a hidden layer of 64 units. This indicates good agreement between observations and predicted values. The value is slightly lower than the minimum adjustment applied to the CARINA dataset ($6 \mu\text{mol/kg}$) for the secondary quality control (QC2) procedures (Tanhua et al., 2010).

A boundary of $2 \times \text{STD}$ (i.e., $10.8 \mu\text{mol kg}^{-1}$ for 3DwMLR and $10.4 \mu\text{mol kg}^{-1}$ for the neural network) has been set as the confidence interval of observations to keep. This way 3462 and 3526 A_T observations (i.e., 4.5% and 4.6%), respectively fall outside this $2 \times \text{STD}$ limit and could then be flagged as questionable or alternatively, replaced by their 3DwMLR or neural network values.

3.4. Data results

A comparison of the data results of both methods has been graphically summarised in Figs. 4 and 5. The first column of both figures displays the difference between A_T obtained by neural network against the 3DwMLR technique. In this column, the full set of 222,136 predicted A_T values is compared. The scale is $\pm 15 \mu\text{mol kg}^{-1}$ and vertical dashed grey lines are placed at -12 , -6 , 0 , 6 and 12 to allow an easy comparison with the ± 6 boundary of minimum adjustment established in CARINA. For each box in the boxplots, the height is proportional to the quantity of samples available between the boxes' boundaries. Boxes have black lines but they were filled in red colour in order to emphasise the intervals of the plotted parameters (presented in left axis) where most of the data were located. The red colour filling is not even perceptible for intervals with less data.

3.4.1. Intercomparison between 3DwMLR and neural network

The top-left histogram of Fig. 4 indicates a good level of agreement between both estimation techniques, giving a standard deviation of $2.36 \mu\text{mol kg}^{-1}$ for the residuals and a mean of $0.074 \mu\text{mol kg}^{-1}$. Thus, 95% of the residuals fall between $\pm 4.7 \mu\text{mol kg}^{-1}$ of A_T .

By analysing the box plots against some other parameters in the first column (Figs. 4 and 5), a good agreement can also be seen. Especially remarkable is the agreement in the estimations of both methods for the same cruises (Cruise ID subplot), indicating that both approaches give similar mean results for each cruise, even when both methods have some disagreements with the measured data, as can be seen in the second and third columns of the CruiseID subplot in Fig. 4. This can be interpreted as an indication of data bias for these cruises. Conversely, poor agreement can be identified as pertaining to samples with low salinity and high oxygen, corresponding to surface samples (shallower than about 200 m) close to coastal areas affected by continental inputs, i.e., in the Labrador Current and Southern Ocean (Fig. 5). Variability in the short time scale is high due to water mixing (Rintoul, 2011; Wootton et al., 2008), such that large residuals are expected for prediction methods.

3.4.2. Comparison of both methods against measured data

The second (the 3DwMLR method) and third (the neural network method) columns of both Figs. 4 and 5 compare estimated A_T vs. observations. Along these columns, only the subset of 76,553 samples with A_T observations can be compared.

After analysing both figures, a good overall agreement between observations and estimated A_T can be stated for both techniques, with some minor variations between them. The neural network technique seems to behave slightly more consistently with geographical parameters, as can be seen in subplots against depth and even with longitude.

Subplots against Cruise Year in Fig. 4 show a slightly reduced scatter for more recent cruises, which is to be expected due to the improvement of analytical techniques and instrumentation. An

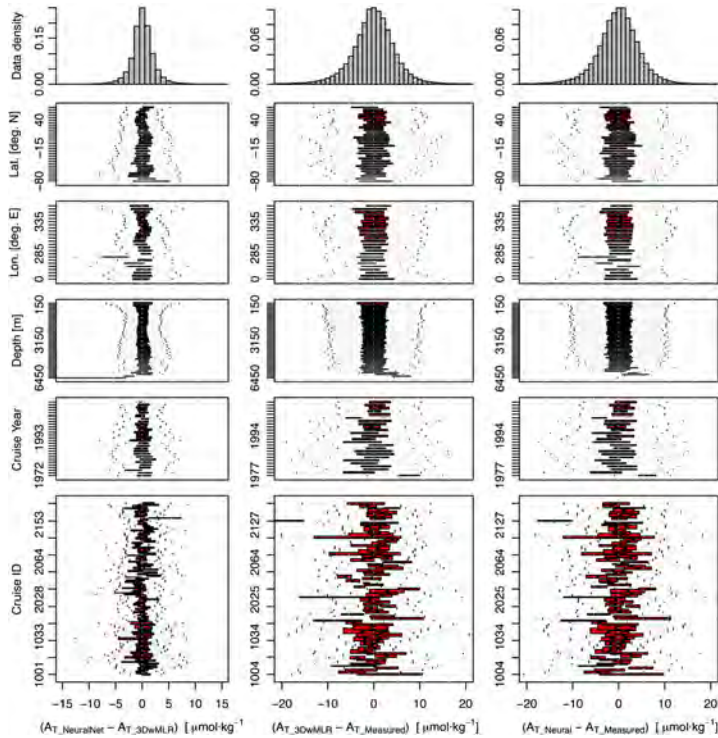


Fig. 4. Histograms and box plots of A_T differences between, from left to right: neural network minus 3DwMLR, 3DwMLR minus observations and neural network minus observations. From top to bottom, first row is histograms of data density, and box-whisker diagrams of the A_T difference against the parameter are indicated on the left: latitude, longitude, depth, year of cruise, and a numerical identifier for each cruise.

abrupt step near 1992 was expected to be seen in these yearly subplots, due to the progressive introduction of CRM in successive years, but it is not appreciable. This is probably due to the use of the GLODAP and CARINA data products that are already adjusted in order to minimise offsets between cruises (Sabine et al., 2005; Tanhua et al., 2009).

As previously indicated, a significant bias against measured A_T data can be stated for some cruises when analysing Cruise ID subplots. Even when the agreement between both estimation techniques themselves is very good for these cruises, it is poor when the comparison is made against the observations of A_T . Cruises with a mean bias greater than $\pm 6 \mu\text{mol}\cdot\text{kg}^{-1}$ are listed in Table 2. All of them correspond to cruises with none or too few deep samples ($> 1500 \text{ m}$) to qualify for secondary quality control and to be adjusted in the CARINA Product (Tanhua et al., 2009, 2010; Velo et al., 2010). As a result of this work, we propose adjusting these five cruises to compensate the bias indicated in Table 2.

By visual inspection of the residuals against seawater properties (Fig. 5) similar patterns are observed for both techniques. The neural network technique seemed to behave more consistently for most parameters, but these differences are not significant.

The highest scatter in Fig. 5 corresponds to low salinity samples for all three columns. Conversely, high oxygen samples show good agreement against observations for both techniques, even though poor agreement between the two estimation methods is indicated in the first column.

4. Conclusions

From an initial dataset with 72,796 samples useful for carbon calculations, i.e., with at least two carbon parameters measured, the algorithms presented here allowed for the extension of the dataset up to 104,043 useful samples. This implies a 42.9% increase in the abundance of carbonate system parameters in the dataset.

Moreover, both proposed methods of estimating alkalinity in the ocean give acceptable results, as the average error in the prediction is lower than $5 \mu\text{mol}\cdot\text{kg}^{-1}$, which can be considered an acceptable uncertainty for alkalinity measurements. Nevertheless, the two approaches to solve the problem are very different and take advantage of different considerations.

Considering only the performance of the two methods used here, measured on how well they can model the natural variability of the

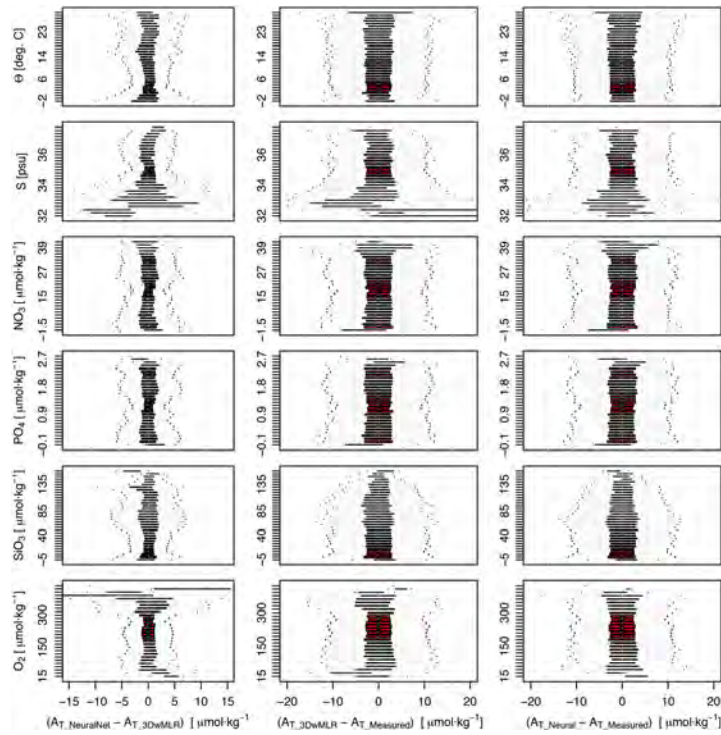


Fig. 5. Box and whisker diagrams of A_T differences between, from left to right: neural network minus 3DwMLR, 3DwMLR minus observations and neural network minus observations. From top to bottom, the parameter being compared against the A_T residuals (indicated on the left): θ , salinity, nitrate, phosphate, silicate and oxygen.

estimated parameter, the neural network performs better, because it is the best at predicting the natural variability (Fig. 3). The results shown in Fig. 3 indicate clearly that the proposed neural network with 64 neurons in the hidden layer and Levenberg–Marquadt tuning (Hagan and Menhaj, 1994) achieves better overall estimations than the 3DwMLR method.

Considering the scalability of the methodology used to obtain the results, the 3DwMLR is the best option, because it delivers a good agreement with the measured data and this methodology can be easily tuned if some misfit is found in some areas. In the 3DwMLR, all the steps performed for estimating every single sample can be

debugged and improved. The neural network represents the very opposite approach. It is true that a tuned network can be retrained, but the weights and the way they are obtained are not transparent, to the point that the combination of the network with its tuning can be considered as a black box that works with the given dataset.

The advantage of using both techniques simultaneously for this purpose is that their results can be easily compared against each other and thus, their performance and agreement assessed.

Acknowledgements

We wish to acknowledge the hard work and dedication of analysts, investigators and crew that collect the data at sea; without them, this work would not be possible. We also gratefully acknowledge those who have contributed their data to the CARINA and GLODAP project. Thanks also to those people in the fisheries group of IIM-CSIC for their support in statistical concerns. This work and in particular T.T. was supported through EU FP7 project CARBOCHANGE “Changes in carbon uptake and emissions by oceans in a changing climate”, which received funding from the European Community’s Seventh Framework Programme under grant agreement no. 264879.

Table 2
Mean biases between 3DwMLR and neural technique against observed A_T data. Units for biases and adjusts are $\mu\text{mol}\cdot\text{kg}^{-1}$. Only values with mean bias greater than $\pm 6 \mu\text{mol}\cdot\text{kg}^{-1}$ are shown.

Carina ID	Expcode	3DwMLR bias	Neural net. bias	Proposed adjust
9	06MT19920509	-7.8 ± 9.0	-7.6 ± 8.6	-7.7
28	06MT20021013	-7.2 ± 19.8	-3.8 ± 15.8	Flag bad
51	29CS19771007	7.6 ± 3.9	6.0 ± 2.6	6.8
57	29CD19860904	-6.9 ± 4.0	-5.1 ± 4.5	-6.0
158	67SL19881117	-19.0 ± 4.8	-14.1 ± 6.0	-16.6

References

- Anderson, D.R., Burnham, K.P., 2002. Avoiding pitfalls when using information-theoretic methods. *J. Wildl. Manag.* 66 (3), 912. <http://dx.doi.org/10.2307/3803155>.
- Anderson, L.A., Sarmiento, J.L., 1994. Redfield ratios of remineralization determined by nutrient data analysis. *Global Biogeochem. Cycles* 8, 65–80. <http://dx.doi.org/10.1029/93GB03318>.
- Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Clarendon Press, Oxford.
- Brewer, P.G., Wong, G.T.F., Bacon, M.P., Spencer, D.W., 1975. An oceanic calcium problem? *Earth Planet. Sci. Lett.* 26 (1), 81–87. [http://dx.doi.org/10.1016/0012-821X\(75\)90179-X](http://dx.doi.org/10.1016/0012-821X(75)90179-X).
- Chen, C.-T., Millero, F.J., 1979. Gradual increase of oceanic CO₂. *Nature* 277 (5693), 205–206. <http://dx.doi.org/10.1038/277205a0>.
- Chen, C.-T., Pytkowicz, R.M., 1979. On the total CO₂-titration alkalinity-oxygen system in the Pacific Ocean. *Nature* 281 (5730), 362–365. <http://dx.doi.org/10.1038/281362a0>.
- Dickson, A.G., 2010. The carbon dioxide system in seawater: equilibrium chemistry and measurements. Guide to Best Practices for Ocean Acidification Research and Data Reporting. Publications Office of the European Union, Luxembourg, pp. 17–52.
- Filznöser, P., Seneels, S., Maronna, R., Van Espen, P.J., 2007. Robust multivariate methods in chemometrics. *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis*. Elsevier, Amsterdam, pp. 681–722.
- Fraga, F., Álvarez-Salgado, X.A., 2005. On the variation of alkalinity during phytoplankton photosynthesis. *Cienc. Mar.* 004, 627–639.
- García, H.E., Locarnini, R.A., Boyer, T.P., Antonov, J.I., Zweng, M.M., Baranova, O.K., Johnson, D.R., 2009. World Ocean Atlas 2009, volume 4: nutrients (phosphate, nitrate, and silicate). In: Levitus, S. (Ed.), NOAA Atlas NESDIS 71. U.S. Government Printing Office, Washington, D.C., p. 398.
- Gardner, M., Dorling, S., 1998. Artificial neural networks (the multilayer perceptron) – a review of applications in the atmospheric sciences. *Atmos. Environ.* 32 (14–15), 2627–2636. [http://dx.doi.org/10.1016/S1352-2310\(97\)00447-0](http://dx.doi.org/10.1016/S1352-2310(97)00447-0).
- Gelder, R., La Roche, J., 2002. Redfield revisited: variability of C:N:P in marine microalgae and its biochemical basis. *Eur. J. Phycol.* 37 (1), 1–17. <http://dx.doi.org/10.1017/S0967026201003456>.
- Green, P.J., 1984. Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives. *J. R. Stat. Soc. B Methodol.* 46 (2), 149–192.
- Hagan, M.T., Menhaj, M.B., 1994. Training feedforward networks with the Marquardt algorithm. *IEEE Trans. Neural Netw.* 5 (6), 989–993.
- Hagan, M.T., Demuth, H.B., Beale, M.H., 1996. Neural Network Design. PWS Publishing, Boston, MA.
- Holland, P.W., Welsch, R.E., 1977. Robust regression using iteratively reweighted least-squares. *Commun. Stat. Theory Methods* 6 (9), 813–827.
- Hubert, M., 2008. High-breakdown robust multivariate methods. *Stat. Sci.* 23 (1), 92–119. <http://dx.doi.org/10.1214/088342307000000087>.
- Ilyina, T., Zeebe, R.E., Maier-Reimer, E., Heinze, C., 2009. Early detection of ocean acidification effects on marine calcification. *Global Biogeochem. Cycles* 23, GB1008. <http://dx.doi.org/10.1029/2008GB003278> (11 pp.).
- Juranek, L.W., Feely, R.A., Peterson, W.T., Alin, S.R., Hales, B., Lee, K., Sabine, C.L., Peterson, J., 2009. A novel method for determination of aragonite saturation state on the continental shelf of central Oregon using multi-parameter relationships with hydrographic data. *Geophys. Res. Lett.* 36. <http://dx.doi.org/10.1029/2009GL040778> (6 pp.).
- Key, R.M., Kozyr, A., Sabine, C.L., Lee, K., Wanninkhof, R., Bullister, J.L., Feely, R.A., Millero, F.J., Mordy, C., Peng, T.H., 2004. A global ocean carbon climatology: results from Global Data Analysis Project (GLODAP). *Global Biogeochem. Cycles* 18, GB4031.
- Key, R.M., Tanhua, T., Olsen, A., Hoppema, M., Jutterström, S., Schirnick, C., van Heuven, S., Kozyr, A., Lin, X., Velo, A., Wallace, D.W.R., Mintrop, L., 2010. The CARINA data synthesis project: introduction and overview. *Earth Syst. Sci. Data* 2 (1), 105–121. <http://dx.doi.org/10.5194/essd-2-105-2010>.
- Kim, T.-W., Lee, K., Feely, R.A., Sabine, C.L., Chen, C.-T.A., Jeong, H.J., Kim, K.Y., 2010. Prediction of Sea of Japan (East Sea) acidification over the past 40 years using a multiparameter regression model. *Global Biogeochem. Cycles* 24. <http://dx.doi.org/10.1029/2009GB003637> (14 pp.).
- Kutner, M.H., Nachtsheim, C.J., Neter, J., Li, W., 2004. Applied Linear Statistical Models. McGraw-Hill, IRWIN, Chicago.
- Lee, K., Tong, L.T., Millero, F.J., Sabine, C.L., Dickson, A.G., Goyet, C., Park, G.H., Wanninkhof, R., Feely, R.A., Key, R.M., 2006. Global relationships of total alkalinity with salinity and temperature in surface waters of the world's oceans. *Geophys. Res. Lett.* 33, L19605.
- Lewis, E., Wallace, D., Allison, L., 1998. Program Developed for CO₂ System Calculations. ORNL/CDIAC-105 Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee.
- Miller, A.J., 1984. Selection of subsets of regression variables. *J. R. Stat. Soc. A Gen* 147 (3), 389–425. <http://dx.doi.org/10.2307/2981576>.
- Millero, F.J., 1995. Thermodynamics of the carbon dioxide system in the oceans. *Geochim. Cosmochim. Acta* 59 (4), 661–677. [http://dx.doi.org/10.1016/0016-7037\(94\)00354-O](http://dx.doi.org/10.1016/0016-7037(94)00354-O).
- Millero, F., 2007. The marine inorganic carbon cycle. *Chem. Rev.* Columbus 107 (2), 308–341.
- Millero, F.J., Lee, K., Roche, M., 1998. Distribution of alkalinity in the surface waters of the major oceans. *Mar. Chem.* 60 (1–2), 111–130. [http://dx.doi.org/10.1016/S0304-4203\(97\)00084-4](http://dx.doi.org/10.1016/S0304-4203(97)00084-4).
- Redfield, A.C., 1934. On the proportions of organic derivatives in sea water and their relation to the composition of plankton. *James Johnstone Memorial Volume*, pp. 176–192.
- Riebesell, U., Fabry, V.J., Hansson, L., Gattuso, J.P., 2010. Guide to Best Practices for Ocean Acidification Research and Data Reporting. Publications Office of the European Union, Luxembourg.
- Rintoul, S.R., 2011. The southern ocean in the earth system. *Science Diplomacy: Antarctica, Science, and the Governance of International Spaces*, pp. 175–188.
- Sabine, C.L., Key, R.M., Kozyr, A., Feely, R.A., Wanninkhof, R., Millero, F.J., Peng, T.S., Bullister, J.L., Lee, K., 2005. Global Ocean Data Analysis Project (GLODAP): Results and data. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee ORNL/CDIAC-145, NDP-083, p. 110.
- Sarle, W.S., 1997. Neural Network FAQ. comp.ai.neural-nets.
- Tanhua, T., Olsen, A., Hoppema, M., Jutterström, S., Schirnick, C., Van Heuven, S., Velo, A., Lin, X., Kozyr, A., Alvarez, M., Bakker, D.C.E., Brown, P., Falck, E., Jeansson, E., Lo Monaco, C., Olafsson, J., Perez, F.F., Pierrot, D., Rios, A.F., Sabine, C.L., Schuster, U., Steinfeldt, R., Stendardo, L., Anderson, L.G., Bates, N.R., Bellerby, R.G.J., Blindheim, J., Bullister, J.L., Gruber, N., Ishii, M., Johannessen, T., Jones, E.P., Köhler, J., Kortzinger, A., Metzl, N., Murata, A., Musielewicz, S., Omar, A.M., Olsson, K.A., de la Paz, M., Pfeil, B., Rey, F., Rhein, M., Sijpevan, I., Tilbrook, B., Wanninkhof, R., Mintrop, L., Wallace, D.W.R., Key, R.M., 2009. CARINA Data Synthesis Project. ORNL/CDIAC-157, NDP-091 Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee, 37831–6335.
- Tanhua, T., van Heuven, S., Key, R.M., Velo, A., Olsen, A., Schirnick, C., 2010. Quality control procedures and methods of the CARINA database. *Earth Syst. Sci. Data* 2 (1), 35–49. <http://dx.doi.org/10.5194/essd-2-35-2010>.
- van Heuven, S., Pierrot, D., Lewis, E., Wallace, D., 2009. MATLAB Program Developed for CO₂ System Calculations. ORNL/CDIAC-105b Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee.
- Velo, A., Vázquez-Rodríguez, M., Padin, X.A., Gilcoto, M., Ríos, A.F., Pérez, F.F., 2010. A multiparametric method of interpolation using WOA05 applied to anthropogenic CO₂ in the Atlantic. *Sci. Mar.* 74 (S1), 21–32. <http://dx.doi.org/10.3989/scimar.2010.74s1021>.
- Wallace, D.W.R., 1995. Monitoring global ocean carbon inventories. Ocean Observing System Development Panel. Texas A&M University, College Station, TX, p. 54.
- Wolf-Gladrow, D.A., Zeebe, R.E., Klaas, C., Kortzinger, A., Dickson, A.G., 2007. Total alkalinity: the explicit conservative expression and its application to biogeochemical processes. *Mar. Chem.* 106 (1–2), 287–300. <http://dx.doi.org/10.1016/j.marchem.2007.01.006>.
- Wootton, J.T., Pfister, C.A., Forester, J.D., 2008. Dynamic patterns and ecological impacts of declining ocean pH in a high-resolution multi-year dataset. *Proc. Natl. Acad. Sci.* 105 (48), 18848–18853. <http://dx.doi.org/10.1073/pnas.0810079105>.
- Zeebe, R.E., Wolf-Gladrow, D.A., 2001. CO₂ in Seawater: Equilibrium, Kinetics, Isotopes. Elsevier Oceanography Series, Amsterdam.

Capítulo IV

An update of anthropogenic CO₂ storage rates in the western South Atlantic basin and the role of Antarctic Bottom Water



Contents lists available at SciVerse ScienceDirect

Journal of Marine Systems

journal homepage: www.elsevier.com/locate/jmarsys



An update of anthropogenic CO₂ storage rates in the western South Atlantic basin and the role of Antarctic Bottom Water

A.F. Ríos ^{a,*}, A. Velo ^a, P.C. Pardo ^a, M. Hoppema ^b, F.F. Pérez ^a
^a Instituto de Investigaciones Marinas, IIM-CSIC, c/Eduardo Cabello 6, E-36208 Vigo, Spain

^b Alfred Wegener Institute for Polar and Marine Research Climate Sciences Postfach 120161, D-27515 Bremerhaven, Germany

ARTICLE INFO

Article history:

Received 21 July 2011

Received in revised form 24 November 2011

Accepted 29 November 2011

Available online 8 December 2011

Keywords:

Anthropogenic CO₂ storage rate

Anthropogenic CO₂ inventory

Decadal variability

Antarctic bottom water

South Atlantic

Southern Ocean

ABSTRACT

The western basin of the South Atlantic from 10°N to 55°S and from the coast to the Mid-Atlantic Ridge is a region with large uncertainties as to the storage of anthropogenic CO₂ (Cant). Our analysis of data of the last three decades provides a Cant storage rate of $0.92 \pm 0.13 \text{ mol m}^{-2} \text{ y}^{-1}$, i.e., 13%–35% higher than previous estimates in this area. The low but significant Cant concentrations ([Cant]) in the large volume of relatively well ventilated Antarctic Bottom Water (AABW) may well be the underlying cause of this higher storage rate. In fact, the significant contribution in terms of Cant of this ventilated AABW that enters the western South Atlantic Ocean was calculated to be $0.055 \pm 0.02 \text{ Pg C y}^{-1}$ or $0.20 \text{ mol m}^{-2} \text{ y}^{-1}$. Instead of being based on the annual trend, the Cant specific inventory (in mol m^{-2}) evolution is more consistently computed as a function of the atmospheric xCO₂ perturbation in ppm, ($0.64 \text{ mol m}^{-2} \text{ ppm}^{-1}$). This methodology allows improved projections of Cant storage rates over long periods.

Published by Elsevier B.V.

1. Introduction

The Atlantic Ocean stores 38% of the oceanic anthropogenic CO₂ (Cant) although it only represents 29% of the global ocean surface area (Sabine et al., 2004). The most ventilated waters are located above the isopycnal of σ_θ 27.2 kg m⁻³, which approximately corresponds to the 5 °C isotherm that separates the large volume of low-Cant cold waters (86% of the Atlantic Ocean volume) from high-Cant warmer upper waters (Vázquez-Rodríguez et al., 2009b). The formation of deep waters in the North Atlantic (North Atlantic Deep Water, NADW) and in the Southern Ocean (Antarctic Bottom Water, AABW) enhances significantly the Cant storage (Lo Monaco et al., 2005; McNeil et al., 2007; Pérez et al., 2008, 2010; Steinfeldt et al., 2009). The Atlantic sector of the Southern Ocean contains moderate concentrations of Cant ([Cant]) in deep waters, i.e., around $10 \mu\text{mol kg}^{-1}$, yet its massive volume turns it into one of the largest carbon reservoirs of the Atlantic basin (Lo Monaco et al., 2005; Vázquez-Rodríguez et al., 2009b). This same significant contribution of Cant ($\sim 10 \mu\text{mol kg}^{-1}$) was found below 4000 m between 30°S and 50°S by Ríos et al. (2010) for the year 1994, while previous estimates for that region showed zero and negative values beneath this depth (Lee et al., 2003; Sabine et al., 2004). The importance of the northwards penetration of the ventilated AABW into the Atlantic basin pointed out by Schlitzer (2007), suggests that a considerable amount of Cant is being transported within

it. McNeil et al. (2007) found considerable uptake of Cant related to the formation of AABW ($0.4 \pm 0.25 \text{ Pg C y}^{-1}$).

We provide an improved estimate of the Cant storage of the western basin of the South Atlantic which in this study runs from 10°N to 55°S and from the coast to the Mid-Atlantic Ridge (about 15°W) (Fig. 1). This region is characterized by important and complex dynamics which involve several circulation cells in upper and intermediate levels; these connect with the western boundary layer and several recirculation features and eastwards escape to the ocean interior in deep and bottom layers (Mémerly et al., 2000). There are some few estimates of CO₂ storage rates in this region (Sabine and Tanhua, 2010). Cant storage along 30°S ($0.6 \pm 0.1 \text{ mol m}^{-2} \text{ y}^{-1}$) was estimated by Murata et al. (2008) for the period 1993–2003. Other estimates in the Atlantic Ocean can be found in the recent studies of Peng and Wanninkhof (2010), who estimated Cant storage rates of 0.53 and $0.36 \text{ mol m}^{-2} \text{ y}^{-1}$ between 15°N and 15°S and 0.83 and $0.35 \text{ mol m}^{-2} \text{ y}^{-1}$ south of 15°S, respectively, and that of Wanninkhof et al. (2010), who calculated Cant storage rates of $0.20 \text{ mol m}^{-2} \text{ y}^{-1}$ between 15°N and 15°S and $0.76 \text{ mol m}^{-2} \text{ y}^{-1}$ south of 15°S.

Since one of the challenges of the carbon community is to obtain reliable regional Cant storages with low uncertainties using repeated sections and high-quality databases, an updated estimation of the Cant storage rate in the western South Atlantic basin is presented here using data from GLODAP (<http://cdiac.ornl.gov/oceans/glodap/>) and CARINA (<http://cdiac.ornl.gov/oceans/CARINA/>) databases covering the period from 1972 to 2005. The contribution of deep and bottom waters in the Cant inventory is furthermore scrutinized. The improvement of carbon measurements in quality and quantity allows

* Corresponding author. Tel.: +34 986 231930x371; fax: +34 986 292762.
E-mail address: aida@iim.csic.es (A.F. Ríos).

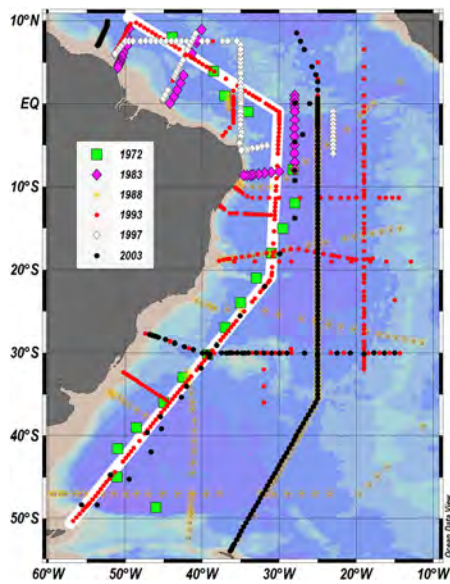


Fig. 1. Map of the western South Atlantic Ocean showing the tracks of the twenty cruises used in this study grouped in 6 reference years that are marked on the map with different symbols. Section WOCE A17 is marked with a white background.

reevaluating the role of the deep water masses, especially the AABW, in the storage rates of Cant showing that their contribution could reach 22% of the total storage rate in the studied zone.

2. Dataset

A total of twenty cruises belonging to the GLODAP and CARINA databases (Hoppema et al., 2009; Key et al., 2010; Tanhua et al., 2009, Tanhua et al., 2010a, 2010b) with high-quality carbon system measurements were selected to study the temporal evolution of the Cant storage in the western South Atlantic basin. Only data from 100 m depth to the bottom were considered in order to avoid the high variability of near-surface measurements. The combined dataset covers 33-years (1972–2005). The cruises were grouped in 6 nominal or reference years (Table 1) to improve the spatial coverage by year. In the earliest cruises total Alkalinity (A_T) and total inorganic Carbon (C_T) were potentiometrically determined without Certified Reference Materials being available to check the consistency of the data. Corrections for TTO-TAS legs 1 and 2 (cruise 46, 1982–3), and SAVE legs 1 and 2 (cruise 48, 1987–8) were taken from Gruber et al. (1996). The A_T offset of $-8 \mu\text{mol kg}^{-1}$ in cruise 24 (Section A17, 1993) reported in Rios et al. (2005) has been also applied. An additional quality control was applied in order to check the A_T and C_T values of the GEOSECS dataset. Carbon measurements in GEOSECS were conducted using the classic potentiometric curves and fitting the pH electrode response to the HCl additions. Most of the uncertainties in these old techniques are due to the precision in the exact volume of sample enclosed in the measuring system. Taking into account that A_T is a steady-state variable that can be very well fitted as a function of salinity and silicate with low uncertainties, we have followed some of the quality control procedures used in GLODAP and CARINA. Using the complete

Table 1

Western South Atlantic cruises (see also Fig. 1) used in this study. Cruises were grouped in six reference years.

Reference year	Cruise year	Cruise code	Carina expo code/ GLODAP CODE	WOCE section or original name
1972	1972	43	GEOSECS_1-	GEOSECS
1983	1982–3	46	TTO-TAS_1-3	TTO-TAS
1988	1987–8	48	318MSAVE_1-5	SAVE
			318MHYDROS	
1993	1991	13	06MT15_3	A09
1993	1991	22	OACES91_1-2	A16S
1993	1993	14	06MT22_5	A10
1993	1993	23	OACES93	A16N
1993	1994	12	06MT19941012	
1993	1994	21	316N142_	A15
1993	1994	24	3230CITHER2_1-2	A17
1993	1995	95	35LU19950909	
1997	1996	84	33LK19960415	
1997	1997	25	316N151_3	A20
1997	1999	106	35TH19990712	
2003	2001	61	29HE20010305	FICARAM
2003	2002	62	29HE20020304	FICARAM
2003	2003	68	316N20030922	
2003	2003	86	33RO20030604	
2003	2003	113	49NZ20031106	Beagle
2003	2005	87	33RO20050111	

GLODAP and CARINA databases, an expected value for each A_T measured was computed. The anomalies between measured and estimated (through Multivariate Linear Regression, MLR) values have a standard deviation (STD) of $\pm 6 \mu\text{mol kg}^{-1}$, which happens to be the same STD as used as lower boundary to flag measured cruise data in the quality control of GLODAP and CARINA. About 25% of GEOSECS measurements, showing deviations higher than $12 \mu\text{mol kg}^{-1}$ (2 STD), were discarded.

3. Results

The main water masses found in the western South Atlantic basin were identified by Mémery et al. (2000) and Rios et al. (2010) on the basis of their hydrographic properties, ventilation and tracers (CFCs). Shown are the salinity (Fig. 2a) and CFC distributions (Fig. 2b) including the isopycnal surfaces which separate the main water masses found in the region. In order to evaluate the variation of [Cant] increase rates and Cant storage rates, the water column was divided into six layers by potential density (σ_θ) intervals following Mémery et al. (2000) and according to the spreading of the different water masses (Table 2). From 100 m depth to $\sigma_\theta < 26.5 \text{ kg m}^{-3}$ the South Atlantic Central Water (SACW) spreads in the subsurface layer of the subtropical waters constituting the SACW layer. Below that, with $26.5 < \sigma_\theta < 27.1 \text{ kg m}^{-3}$, the Sub Antarctic Mode Water (SAMW) is found in the subpolar region. The Antarctic Intermediate Water (AAIW) spreads in the density range of $27.1 < \sigma_\theta < 27.4 \text{ kg m}^{-3}$. In the density range delimited by $\sigma_\theta > 27.4 \text{ kg m}^{-3}$ and $\sigma_\theta < 41.47 \text{ kg m}^{-3}$ the upper North Atlantic Deep Water (uNADW) and the upper Circumpolar Deep Water (uCDW) can be found mainly in the north and south of the region, respectively. Both lower North Atlantic Deep Water (lNADW) and lower Circumpolar Deep Water (lCDW) constitute the lNADW-lCDW layer and spread in the density range of $\sigma_\theta > 41.47$ and $\sigma_\theta < 45.90 \text{ kg m}^{-3}$. Finally, the AABW can be found throughout $\sigma_\theta > 45.90 \text{ kg m}^{-3}$.

Estimates of [Cant] were obtained by applying the φC_T^2 method (Vázquez-Rodríguez et al., 2009a, 2009b) whose uncertainty is $\pm 5.2 \mu\text{mol kg}^{-1}$, and the results are shown in Fig. 2c. This method is based on the ΔC_T^2 principles (Gruber et al., 1996) with some improvements. In the φC_T^2 method the subsurface layer is used as the reference to compute the A_T and air–sea CO_2 disequilibrium since this layer is able to preserve the conditions of water mass formation at an annual scale. Furthermore and importantly, these two terms, different to all other back-calculation methods, are assumed to be in

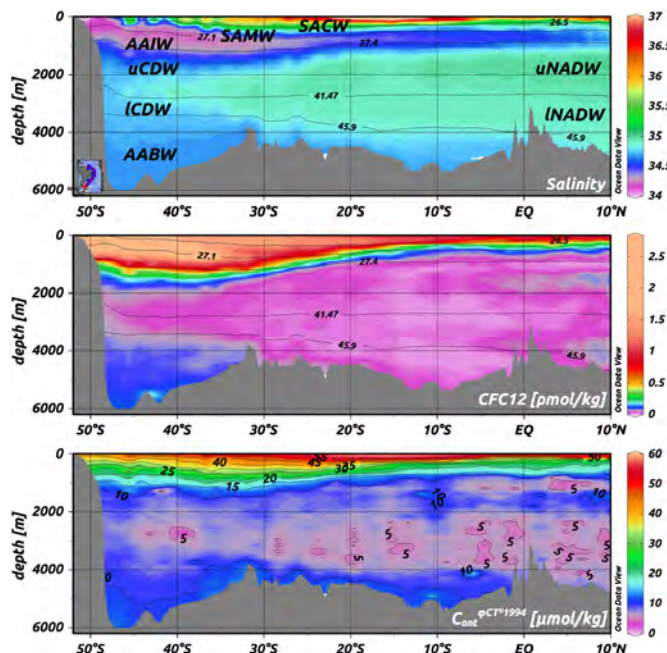


Fig. 2. Vertical sections of salinity, CFC12 (pmol kg^{-1}) and C_{ant} ($\mu\text{mol kg}^{-1}$) along the WOCE A17 line in 1994 (see Fig. 1). The selected isopycnals (potential density “ σ_t ”, in kg m^{-3}) used to separate and follow the C_{ant} evolution of the main water masses are also depicted. Water mass acronyms: SACW = Sub Atlantic Central Water, SAMW = Sub Antarctic Mode Water, AAIW = Antarctic Intermediate Water, NADW = North Atlantic Deep Water, CDW = Circumpolar Deep Water, AABW = Antarctic Bottom Water. The lowercase first letter “u” and “l” denotes the “upper” and “lower” varieties in some water masses.

non-steady-state since the pre-industrial era (see Appendix A for more details of the method). The highest [Cant] is found in shallow waters ($\sigma_t < 27.4$, Fig. 2c) since Cant is introduced into the ocean when the water masses are in contact with the atmosphere. Down to intermediate layers of the water column, [Cant] reduces progressively and so does the level of ventilation (Fig. 2b and c). However, in deep and bottom waters, significant values of [Cant] occur with a maximum at the bottom accompanied by a slight increase in the level of ventilation compared to that of intermediate layers (Fig. 2b and c).

Table 2
Potential density intervals (σ_t) for the six layers where the main water masses of the region spread. Layer thickness in meters \pm the standard deviation. Nominal depths of the isopycnals in parentheses.

Layers	Layer thickness (m)	Potential density intervals (kg m^{-3})
SACW	168 \pm 32	100 m to $\sigma_t < 26.5$ (300 m)
SAMW	306 \pm 25	26.5 $< \sigma_t < 27.1$ (600 m)
AAIW	412 \pm 14	27.1 $< \sigma_t < 27.4$ (1100 m)
uNADW-uCDW	1594 \pm 68	$\sigma_t > 27.4$ and $\sigma_t < 41.47$ (3000 m)
INADW-ICDW	532 \pm 166	$\sigma_t > 41.47$ and $\sigma_t < 45.90$ (4000 m)
AABW	902 \pm 311	$\sigma_t > 45.90$ (down to bottom)

The Cant specific inventory ($Inv_{\text{Cant}}^{\text{c}}$), in mol m^{-2} , was calculated for the study region following the methodology of Pérez et al. (2010), and is defined as:

$$Inv_{\text{Cant}}^{\text{c}} = \sum_{l=1}^6 \rho_{l,c} \cdot p_t \cdot C_{\text{ant}}^{l,c} \cdot Th_{l,c}$$

where $C_{\text{ant}}^{l,c}$ stands for the [Cant] average for the set of cruises “c” in the layer “l” of the water column (in $\mu\text{mol kg}^{-1}$), ρ is the seawater density (in kg m^{-3}) which was taken from the WOA09 climatology (2009) and $Th_{l,c}$ is the thickness of the layer “l” of the water column (in meters, see Table 2).

The spatial coverage of measurements during each reference year (Table 1, Fig. 1) is variable and, therefore the properties of the water masses in terms of tracers can cause significant differences between the observed average layer properties in each reference year. Such differences are likely to introduce potential spatial biases in the $Inv_{\text{Cant}}^{\text{c}}$ estimates that must be corrected. In order to apply this correction, [Cant] was estimated in each layer through a MLR analysis with salinity, potential temperature, AOU (Apparent Oxygen Utilization), nutrients, and year as independent variables (Pérez et al., 2010). Each one of the coefficients of the MLR were multiplied by the difference between the averaged value of these variables in each layer (all

data obtained from WOA09 database) and the averaged values of the variables considering only those points in the WOA09 coinciding with the positions of the cruises. Accordingly, the $C_{\text{ant}}^{\text{LC}}$ computed initially was corrected by adding a new term $\Delta C_{\text{ant}}^{\text{LC}}$. The range of $\Delta C_{\text{ant}}^{\text{LC}}$ varied between -1 and $+1.2 \mu\text{mol kg}^{-1}$.

The temporal evolution of [Cant] averages ($C_{\text{ant}}^{\text{LC}}$) in each of the six layers for the period 1972–2003 is plotted in Fig. 3a together with the temporal evolution of the $\text{Inv}_{\text{Cant}}^{\text{LC}}$ for the western South Atlantic basin. The error bars represent $\pm 2\sigma/N^{0.5}$ (where σ is the standard deviation and N the number of used data). The uncertainty is relatively low because we have gathered a large number of quality controlled data (CARINA and GLODAP). The [Cant] averages decreased with depth except for the AABW layer, where higher [Cant] average values than those in the INADW–ICDW and uNADW–uCDW layers were found. From the temporal evolution of [Cant] averages the different [Cant] increase rates are obtained in each layer. The upper SACW layer shows a [Cant] rate of increase of $0.90 \pm 0.04 \mu\text{mol kg}^{-1} \text{ y}^{-1}$ following atmospheric CO_2 evolution at a level very close to the CO_2 saturation ($\sim 90\%$). In the SAMW layer, the [Cant] increase rate ($0.53 \pm 0.02 \mu\text{mol kg}^{-1} \text{ y}^{-1}$) is lower than that obtained in the upper layer. Regarding the AAIW layer, the [Cant] increase rate ($0.36 \pm 0.02 \mu\text{mol kg}^{-1} \text{ y}^{-1}$) is about 50% lower than that in the SAMW layer. From these intermediate layers down a sharp decrease of the [Cant] increase rates is found. However, in the AABW layer the [Cant] increase rate found ($0.15 \pm 0.04 \mu\text{mol kg}^{-1} \text{ y}^{-1}$) is higher than that of the INADW–ICDW layer and of the same order of that in the uNADW–uCDW layer.

The $\text{Inv}_{\text{Cant}}^{\text{LC}}$ showed an increase over 1972–2003 with a storage of $0.92 \pm 0.13 \text{ mol m}^{-2} \text{ y}^{-1}$. In order to estimate the Cant inventory in the western South Atlantic basin, the area from 10°N to 55°S was taken to be $22.64 \cdot 10^{12} \text{ m}^2$. The Cant inventory was $14 \pm 0.8 \text{ Pg C}$ for the year 1994 and $16.5 \pm 0.8 \text{ Pg C}$ for the year 2004 which corresponds to an average Cant storage of $0.250 \pm 0.035 \text{ Pg C y}^{-1}$.

Usually, [Cant] increase rates and Cant storage rates are obtained by linear regression of [Cant] and Cant specific inventories, respectively, versus year assuming a linear relationship (Murata et al., 2008; Peng and Wanninkhof, 2010; Pérez et al., 2008, 2010; Rios et al., 2001; Wanninkhof et al., 2010). This is quite convenient for relatively short periods of time since it permits to detect changes in [Cant] beyond their uncertainties. To detect changes in deep waters with low [Cant], long periods of time should be considered because of the small signal of [Cant]. However, during long periods, the rate of increase of the atmospheric CO_2 is significantly accelerating and thus, the [Cant] of the water masses equilibrated totally or partially

should follow an exponential rate of increase (Steinfeldt et al., 2009; Tanhua et al., 2007). Therefore, maintaining a linear adjustment would result in biased estimated [Cant] increase rates and Cant storage rates.

On the other hand, adjusting the variables to an exponential curve would result in unreliable values. A way to improve Cant storage rate and [Cant] increase rate estimations is to fit [Cant] of the water masses and the Cant specific inventories to a linear relationship with the excess of atmospheric xCO_2 above the pre-industrial value of $280 \mu\text{mol mol}^{-1}$ (ΔxCO_2) (Fig. 3b). Here we assume the ocean circulation to be in steady state and that a conservative tracer, with exponentially increasing concentration in the surface, reaches a 'transient steady state' after a time significantly longer than the surface growth timescale, as Gammon et al. (1982) demonstrated for a one-dimensional system with no lateral advection but only diffusion and vertical advection. Tanhua et al. (2006) found that Cant is in 'transient steady state' even though the Cant is supplied by lateral advection of recently ventilated waters. Therefore, the Cant concentration increases proportionally over time through the whole water column which is directly related to its time-dependent surface concentration. The asset of this fit is based on the non-existence of an intercept, reducing the degrees of freedom. In fact, all the layers have zero intercept except for the SAMW layer. These modified fits have more significant statistical parameters and reduce the slope uncertainties. Since [Cant] in surface waters, assuming CO_2 -saturation, follows an e-folding time function with a rate of increase of 1.69 y^{-1} (Steinfeldt et al., 2009; Tanhua et al., 2007), ΔxCO_2 increases in the same way and the Cant inventory can be computed as 0.0169 times the Cant storage. Thus with this increase rate (1.69 y^{-1}) and the Cant storage rate ($0.64 \pm 0.13 \text{ mol m}^{-2} \text{ ppm}^{-1}$), a Cant storage rate based on ΔxCO_2 of $10.8 \pm 2.2 \text{ mmol m}^{-2} \text{ y}^{-1} \text{ ppm}^{-1}$ is obtained, which corresponds to $0.53 \pm 0.11 \text{ mol m}^{-2} \text{ y}^{-1}$ ($0.144 \pm 0.030 \text{ Pg-C y}^{-1}$) and $1.05 \pm 0.21 \text{ mol m}^{-2} \text{ y}^{-1}$ ($0.285 \pm 0.057 \text{ Pg-C y}^{-1}$) for the years 1972 and 2003, respectively (note that ΔxCO_2 values for 1972 and 2003 are 31.33 and $62.28 \mu\text{mol mol}^{-1}$, respectively). The Cant storage rates in $\text{mol m}^{-2} \text{ y}^{-1}$ calculated for each year are also represented in Fig. 3. These Cant storage rates are multiplied by 1.69 y^{-1} so the equivalence with the Cant specific inventories is more clearly shown. Both trends are in good agreement, showing the Cant specific inventories calculated by year from ΔxCO_2 slightly lower values in 1972 and higher in 2003 than the Cant specific inventories previously estimated for each nominal year. Importantly, the improvement in the estimation of the [Cant] increase rates through this method with

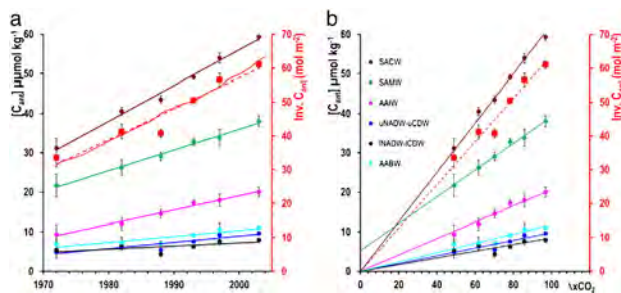


Fig. 3. Evolution of [Cant] averages in $\mu\text{mol kg}^{-1}$ (dots) and Cant specific inventories in mol C m^{-2} (squares) versus year (a) and versus the excess of atmospheric xCO_2 (ΔxCO_2) (b), in each layer of water mass (legend in plot b). The error bars represent $\pm 2\sigma/N^{0.5}$ (where σ is the standard deviation and N the number of used data). The [Cant] rates of increase (continuous line) in $\mu\text{mol kg}^{-1} \text{ y}^{-1}$ (a) and $\mu\text{mol kg}^{-1} \text{ ppm}^{-1}$ (b), and the Cant storage rates (dashed line) in $\text{mol m}^{-2} \text{ y}^{-1}$ (a) and in $\text{mol m}^{-2} \text{ ppm}^{-1}$ (b) are given in Table 3. The Cant storage rates in $\text{mol m}^{-2} \text{ y}^{-1}$ (thin line) was calculated from the Cant storage rate ($0.64 \text{ mol m}^{-2} \text{ ppm}^{-1}$) at each year according to the ΔxCO_2 , and represented multiplied by 0.0169 (right axis).

respect to the linear temporal trends is very significant in deep and bottom layers, which altogether account for approximately 45% of the Cant inventories in the water column.

4. Discussion

As Wanninkhof et al. (2010) pointed out, decadal changes in deep water are more subtle than in subsurface and intermediate waters. Nevertheless, in terms of Cant inventories, the detection of changes at depths below 2000 m are of relevance by virtue of the large volume of water involved (Garzoli et al., 2010); indeed, in the Atlantic Ocean, cold waters ($<5^{\circ}\text{C}$, i.e., below $\sim 900\text{ m}$) represent 86% of its total volume (Vázquez-Rodríguez et al., 2009b). Although the penetration of Cant below 2000 m was suggested to be small in some reconstructions of Cant in the Atlantic (Gruber, 1998; Lee et al., 2003; Sabine et al., 2004), it is noticeable at southern latitudes (Lo Monaco et al., 2005; Rios et al., 2010; Vázquez-Rodríguez et al., 2009b). This increasing tendency of Cant specific inventories south of 30°S is due to the use of new methods of Cant estimation instead of the ΔC method. The ΔC method showed zero and even negative Cant concentrations beneath 4000 m (Lee et al., 2003; Sabine et al., 2004). These relatively high Cant concentrations in the southern Atlantic deep waters are due to the penetration of the ventilated AABW from the Southern Ocean (Schlitzer, 2007). This can also be seen in the distribution of CFC-12 along the A17 section (Fig. 2b). Therefore, bottom waters should not be neglected in computing the Cant inventory.

The Cant inventory of $14 \pm 0.8\text{ Pg C}$ here computed (referred to 1994) is very similar to that obtained by Waugh et al. (2006) (14.1 Pg C for the west South Atlantic) using the Transit Time Distribution (TTD) method that commonly uses CFC-11 or CFC-12 concentrations as proxies of the anthropogenic CO_2 signal, though the TTD estimate could be biased due to the assumption of constant disequilibrium. However, the Cant inventory computed in this study is around 30% higher than the 10.8 Pg C (referred also to 1994) obtained by Lee et al. (2003) using a back-calculating technique.

The Cant storage rate obtained here for the western South Atlantic basin ($0.92 \pm 0.13\text{ mol m}^{-2}\text{ y}^{-1}$) is higher than that obtained by other authors (Table 4). Cant storage rates in the South Atlantic were previously estimated using isopycnal surfaces (Murata et al., 2008) MLR (Peng and Wanninkhof, 2010) and eMLR (extended MLR) methods (Peng and Wanninkhof, 2010; Wanninkhof et al., 2010). The Cant storage rates obtained by these authors, once rescaled to the western South Atlantic basin (10°N to 55°S), result in significantly lower values than the one obtained in the present study (Table 4). It is important to note that when the Cant storage rate here obtained is only considered for the upper and the intermediate waters down to $\sim 1100\text{ m}$ ($0.47\text{ mol m}^{-2}\text{ y}^{-1}$), the value would be significantly reduced, indicating the importance of the contribution of the deep to bottom layers (Table 3).

To evaluate the C_{ANT} storage rate associated with the bottom waters, mainly AABW, entering in the western South Atlantic Ocean, a schema of the Cant balance in the Southern Ocean is shown in Fig. 4. According to Khatiwala et al. (2009) the Southern Ocean has

an extraordinary contribution of about 40% of the global Cant air-sea uptake. A significant part of this Cant uptake is stored in the Southern Ocean; another part is exported northwards through upper levels as mode and intermediate waters, and still another part through deep levels as AABW and the CDW branches. The ventilation of the deep ocean was evaluated by inventories of CFC-11 in the Southern Ocean by Orsi et al. (2002). They find that the total CFC-11 accumulated in the Atlantic Sector ($16.5 \times 10^6\text{ mol}$) represents about of 55% of the total CFC-11 accumulated ($29.7 \times 10^6\text{ mol}$) in the Southern Ocean in deep water (uLDW, ICDW and AABW). Sabine et al. (2004) estimate from the GLODAP data base that the total Cant inventory in the Southern Ocean is $10 \pm 1.5\text{ Pg C}$ with a small contribution of the Atlantic Sector (2 Pg C) which is about 20% and rather low compared with the CFC-11 inventory. Recently a new computation of Cant inventory of $8.7 \pm 0.7\text{ Pg C}$ has been obtained using the CARINA database (Vázquez-Rodríguez et al., 2009b; Velo et al., 2010) for the South Atlantic south of 45°S . Using the new evaluation for the Atlantic Sector, the total Cant inventory for the Southern Ocean rises to $16.7 \pm 1.7\text{ Pg C}$ and the Atlantic sector would have a contribution of 52% which appears more consistent with the CFC-11 inventories. Applying the annual increase rate of 1.69 y^{-1} set by Steinfeldt et al. (2009), the Southern Ocean would have a Cant storage rate of $0.28 \pm 0.03\text{ Pg C y}^{-1}$ (Fig. 4).

Orsi et al. (2002) also evaluated the thermohaline circulation in the Southern Ocean, reporting that about 21 Sv ($10^6\text{ m}^3\text{ s}^{-1}$) of water is transported northward as ventilated (AABW) and old waters (ICDW, uCDW) through the deep ocean (from $\sim 1500\text{ m}$ to the bottom). If an average [Cant] of about $12\text{ }\mu\text{mol kg}^{-1}$ is assigned to deep water (Lo Monaco et al., 2005; Vázquez-Rodríguez et al., 2009b), the deep northward export (AABW and CDW) would be 0.10 Pg C y^{-1} (Fig. 4). If the northward advection of the upper and intermediate layers (SAMW and AAIW) ranged between 15 and 20 Sv (Hartín et al., 2011) and this is combined with an average [Cant] of about $30\text{ }\mu\text{mol kg}^{-1}$ (Key et al., 2004; Lo Monaco et al., 2005; Vázquez-Rodríguez et al., 2009b), an upper layer export of around 0.25 Pg C y^{-1} is obtained (Fig. 4). All together the total Cant export amounts to 0.35 Pg C y^{-1} . Adding the above Southern Ocean reevaluated total Cant storage (0.28 Pg C y^{-1}), a total air-sea Cant uptake of 0.63 Pg C y^{-1} would be needed to compensate. In spite of very different combined estimations of [Cant] and water flows, this value is only 14% lower than the Cant air-sea uptake (0.73 Pg C y^{-1}) given by Khatiwala et al. (2009), if rescaled for 1994; this could suggest that the estimations of export and storage were a bit low. Assuming that the same ratio of 55% of total CFC-11 of the Southern Ocean (Orsi et al., 2002) derives from the Atlantic sector also holds for the Cant transport in the Atlantic in deep waters, the Southern Ocean deep waters would feed the South Atlantic with a rate of 0.055 Pg C y^{-1} , or $0.20\text{ mol m}^{-2}\text{ y}^{-1}$, predominantly driven by AABW. This contribution represents 22% of the Cant storage in the South Atlantic, leaving enough room for other contributions from the North Atlantic. This strongly suggests that the Southern Ocean deep convection has an important contribution to the uptake and the accumulation of Cant at a global scale.

Table 3

[Cant] increase rates in $\mu\text{mol kg}^{-1}\text{ y}^{-1}$ and $\mu\text{mol kg}^{-1}\text{ ppm}^{-1}$ in the different layers of the water masses. The errors represent $\pm 2\sigma/\text{N}^{0.5}$. Cant storage rates in $\text{mol m}^{-2}\text{ y}^{-1}$ and $\text{mol m}^{-2}\text{ ppm}^{-1}$ for the whole water column of the western South Atlantic (see Fig. 3).

Water mass	Rate $\mu\text{mol kg}^{-1}\text{ y}^{-1}$	Correlation r^2	p-level	Rate $\mu\text{mol kg}^{-1}\text{ ppm}^{-1}$	Correlation r^2	p-level
SACW	0.90 ± 0.04	0.991	0.002	0.627 ± 0.006	0.9996	1E-10
SAMW	0.53 ± 0.02	0.991	0.002	0.34 ± 0.015	0.991	2E-5
AAIW	0.36 ± 0.02	0.985	0.004	0.207 ± 0.004	0.998	5E-7
uNADW-uCDW	0.16 ± 0.04	0.82	0.052	0.098 ± 0.004	0.99	2E-6
INADW-ICDW	0.08 ± 0.04	0.45	0.250	0.085 ± 0.006	0.97	3E-5
AABW	0.15 ± 0.04	0.80	0.061	0.118 ± 0.004	0.99	12E-7
$\text{mol m}^{-2}\text{ y}^{-1}$				$\text{mol m}^{-2}\text{ ppm}^{-1}$		
Cant storage rates for the whole water column	0.92 ± 0.13	0.93	0.019	0.64 ± 0.13	0.997	7E-8

Table 4
Comparison of Cant storages ($\text{mol m}^{-2} \text{y}^{-1}$) for the western South Atlantic basin (10°N to 55°S).

Author	Method	Rates of storage ($\text{mol m}^{-2} \text{y}^{-1}$)
Murata et al. (2008)	Isopycnal	0.80
Peng and Wanninkhof (2010)	MLR (eMLR)	0.74 ± 0.3 (0.35) ¹
Wanninkhof et al. (2010)	eMLRdens	0.60
This study	Backcalculation (φCT^0)	0.92 ± 0.13

¹value in parentheses refers to the eMLR storage rate.

5. Conclusions

The Cant storage rate observed for the western South Atlantic basin ($0.92 \pm 0.13 \text{ mol m}^{-2} \text{y}^{-1}$) is higher than previous estimates. A closer inspection of these differences suggests the low but significant [Cant] in the high volume of AABW that enters the South Atlantic as the reason of the higher Cant storage rate. The Cant storage rate associated to the AABW that enters the western South Atlantic Ocean was calculated to be $0.055 \pm 0.02 \text{ Pg C y}^{-1}$ ($0.20 \text{ mol m}^{-2} \text{y}^{-1}$) which represents 22% of the Cant storage rate observed in our study. The deviations in the Cant storage rate estimates observed between previously used methods (isopycnal surfaces, MLR and eMLR) and the backcalculation φCT^0 method are due to the penetration of deep and bottom waters with low [Cant] from the Southern Ocean and the North Atlantic. This low [Cant] are almost undetectable, mainly by MLR methods, when using data from cruises carried out less than two decades apart. Furthermore, because the exponential atmospheric CO_2 increase can affect the linear long-term trend estimation, a more reliable Cant storage rate was computed as a function of the excess of atmospheric xCO_2 increase ($0.64 \pm 0.13 \text{ mol m}^{-2} \text{ppm}^{-1}$). This storage rate is also very useful in terms of future projections of Cant inventories according to the different scenarios of atmospheric xCO_2 due to the linearity between both terms. The analysis done here for the western South Atlantic Ocean highlights the need to continue with programs of repeated sections with high accuracy carbon measurements (e.g., Van Heuven et al., 2011). This kind of studies allows better detection of the variations of Cant in the voluminous deep and bottom waters, which are of high relevance for obtaining more reliable Cant storage rates.

Acknowledgements

This work was supported through EU FP7 project CARBOCHANGE “Changes in carbon uptake and emissions by oceans in a changing

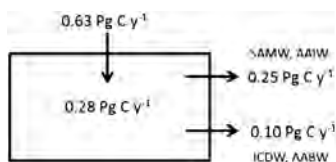


Fig. 4. Schematic of Cant balance in the Southern Ocean. The storage rate of Cant in the box was obtained by rescaling the value of Sabine et al. (2004) modified according to new estimates in the SO-Atlantic sector in agreement with the CFC inventories given by Orsi et al. (2002). The Cant rates of the northward export through the upper and intermediate layers (SAMW and AAIW) was estimated from the 15–20 Sv water transport (Hartin et al., 2011) and $30 \mu\text{mol kg}^{-1}$ for Cant (Key et al., 2004; Lo Monaco et al., 2005; Vázquez-Rodríguez et al., 2009b). The Cant rates of the northward export through the deep waters (ICDW and AABW) was estimated from the 21 Sv water transport (Orsi et al., 2002) and $12 \mu\text{mol kg}^{-1}$ for Cant (Lo Monaco et al., 2005; Vázquez-Rodríguez et al., 2009b). The Cant air-sea uptake was obtained by difference to be in balance with the Cant storage rates and export.

climate" which received funding from the European Community's Seventh Framework Programme under grant agreement no. 264879. We would like to thank the teams who contributed in putting together the GLODAP and CARINA databases and the chief scientists, scientists and crew who participated and put their efforts in the oceanographic cruises used in this study. We would also like to thank Toste Tanhua, Rik Wanninkhof and the anonymous reviewer for their thoughtful comments. Our special gratitude goes to Michel Arhan (Coordinator of the WOCE-France program CITER), who shared his knowledge on the South Atlantic Ocean.

Appendix A

The φCT^0 method to estimate Cant in the Atlantic shares similar fundamentals as the ΔC^* back-calculation method (Gruber et al., 1996). The sub-surface layer (100–200 m) is taken in the φCT^0 method as a reference for characterizing water mass properties at the moment of their formation. The air-sea CO_2 disequilibrium ($\Delta\text{C}_{\text{dis}}$) is parameterized at the sub-surface layer first using a short-cut method (Thomas and Ltekkot, 2001) to estimate C_{ANT} . Since the average age of the water masses in the 100–200 m depth domain, and most importantly in outcropping regions, is under 25 years, the use of the short-cut method to estimate C_{ANT} is appropriate (Matear et al., 2003). The A_τ^0 and $\Delta\text{C}_{\text{dis}}$ parameterizations (in terms of conservative tracers) obtained from sub-surface data are applied directly to calculate C_{ANT} in the water column for waters above the 5°C isotherm and via an OMP (optimum multiparametric) analysis for waters with $\theta < 5^\circ\text{C}$. This procedure especially improves the estimates in cold deep waters that are subject to strong and complex mixing processes between Arctic and Antarctic water masses. One important feature of the φCT^0 method is that none of the A_τ^0 or $\Delta\text{C}_{\text{dis}}$ parameterizations are CFC-reliant. Besides that, the φCT^0 approach proposes an approximation to the temporal and spatial variability of $\Delta\text{C}_{\text{dis}}$ ($\Delta\Delta\text{C}_{\text{dis}}$) in the Atlantic Ocean in terms of C_{ANT} and $\Delta\text{C}_{\text{dis}}$ itself. The φCT^0 method expression for the calculation of C_{ANT} is as follows:

$$\text{C}_{\text{ANT}} = \frac{\Delta\text{C}^* - \Delta\text{C}_{\text{dis}}^{\text{CFC}}}{1 + \phi|\Delta\text{C}_{\text{dis}}^{\text{CFC}}/\text{C}_{\text{ANT}}^{\text{sat}}} \quad (1)$$

The ΔC^* is defined after Gruber et al. (1996) as:

$$\Delta\text{C}^* = \text{C}_\tau - \text{AOU}/\text{R}_\tau - 0.5(\text{PA}_\tau - \text{PA}_\tau^0) - \text{C}_\tau^{\text{p}} \text{eq} \quad (2)$$

The constant term “ φ ” is a proportionality factor that stands for the $\Delta\Delta\text{C}_{\text{dis}}/\Delta\text{C}_{\text{dis}}$ ratio and its value (0.55). The $\Delta\text{C}_{\text{dis}}^{\text{CFC}}$ and PA_τ^0 terms are parameterized as a function of conservative parameters exclusively. The $\text{C}_{\text{ANT}}^{\text{sat}}$ stands for the theoretical C_{ANT} saturation concentration depending on the pCO_2 at the time of water mass formation and is defined as $\text{C}_{\text{ANT}}^{\text{sat}} = \text{S}/35 \cdot (0.85 - \theta + 46.0)$ (at present xCO_2 air). Based on earlier uncertainty and error evaluations (Gruber et al., 1996; Lee et al., 2003; Sabine et al., 1999; Touratier et al., 2007), a random propagation of the errors associated with the input variables necessary to solve Eqs. (1 and 2) has been performed and an estimated overall uncertainty of $\pm 5.2 \mu\text{mol kg}^{-1}$ is obtained for the φCT^0 method. For more details: <http://www.biogeosciences-discuss.net/6/4527/2009/bgd-6-4527-2009-print.pdf>

References

- Gammon, R.H., Cline, J., Wisegarver, D., 1982. Chlorofluoromethanes in the northeast Pacific Ocean: measured vertical distributions and application as transient tracers of upper ocean mixing. *J. Geophys. Res.* 87, 9441–9454.
- Garzoli, S.L., Boebel, O., Bryden, H., Fine, R.A., Fukasawa, M., Gladyshev, S., Johnson, G., Johnson, M., MacDonald, A., Meinen, C., Mercier, H., Orsi, A., Piola, A., Rintoul, S., Speich, S., Visbeck, M., Wanninkhof, R., 2010. Progressing towards global sustained deep ocean observations. In: Hall, J., et al. (Eds.), *OceanObs'09: Sustained Ocean Observations and Information for Society*. ESA Publ (WPP-306, 12 pp.).

- Gruber, N., 1998. Anthropogenic CO₂ in the Atlantic Ocean. *Global Biogeochem. Cycles* 12, 165–191.
- Gruber, N., Sarmiento, J.L., Stocker, T.F., 1996. An improved method for detecting anthropogenic CO₂ in the oceans. *Global Biogeochem. Cycles* 10, 809–837.
- Hartin, C.A., Fine, R.A., Sloyan, B.M., Talley, L.D., Chereskin, T.K., Hapell, J., 2011. Formation rates of Subantarctic mode water and Antarctic intermediate water within the South Pacific. *Deep-Sea Research* 58, 524–534.
- Hoppema, M., Velo, A., Van Heuven, S., Tanhua, T., Key, R.M., Lin, X., Bakker, D.C.E., Pérez, F.F., Rios, A.F., Lo Monaco, C., Sabine, C.L., Álvarez, M., Bellerby, R.G.J., 2009. Consistency of cruise data of the CARINA database in the Atlantic sector of the Southern Ocean. *Earth Syst. Sci. Data* 1, 63–75. doi:10.5194/essd-1-63-2009.
- Key, R.M., Kozyr, A., Sabine, C.L., Lee, K., Wanninkhof, R., et al., 2004. A global ocean carbon climatology: results from Global Data Analysis Project (GLODAP). *Global Biogeochemical Cycles* 18, GB4031. doi:10.1029/2004GB002247.
- Key, R.M., Tanhua, T., Olsen, A., Hoppema, M., Jutterström, S., Schirnack, C., Van Heuven, S., Kozyr, A., Lin, X., Velo, A., Wallace, D.W.R., Mintrop, L., 2010. The CARINA data synthesis project: introduction and overview. *Earth System Science Data* 2, 105–121. doi:10.5194/essd-2-105-2010.
- Khatiwala, S., Primeau, F., Hall, T., 2009. Reconstruction of the history of anthropogenic CO₂ concentrations in the ocean. *Nature* 462, 346–349. doi:10.1038/nature08526.
- Lee, K., Choi, S.-D., Park, G.-H., Wanninkhof, R., Peng, T.-H., Key, R.M., Sabine, C.L., Feely, R.A., Bullister, J.L., Millero, F.J., Kozyr, A., 2003. An updated anthropogenic CO₂ inventory in the Atlantic Ocean. *Global Biogeochem. Cycles* 17 (4), 1116. doi:10.1029/2003GB002067.
- Lo Monaco, C., Metzl, N., Poisson, A., Brunet, C., Schauer, B., 2005. Anthropogenic CO₂ in the Southern Ocean: distribution and inventory at the Indian-Atlantic boundary (World Ocean Circulation Experiment line I6). *J. Geophys. Res.* 110, C06010. doi:10.1029/2004JC002643.
- Matear, R.J., Wong, C.S., Xie, L., 2003. Can CFCs be used to determine anthropogenic CO₂? *Global Biogeochem. Cycles* 17, 1013. doi:10.1029/2001GB001415.
- McNeil, B.J., Metzl, N., Key, R.M., Matear, R.J., Corbiere, A., 2007. An empirical estimate of the Southern Ocean air-sea CO₂ flux. *Global Biogeochem. Cycles* 21, GB3011. doi:10.1029/2007GB002991.
- Mémery, L., Arhan, M., Alvarez-Salgado, X.A., Messias, M.-J., Mercier, H., Castro, G.C., Rios, A.F., 2000. The water masses along the western boundary of the south and equatorial Atlantic. *Prog. Oceanogr.* 47, 69–98.
- Murata, A., Kumamoto, Y., Sasaki, K., Watanabe, S., Fukasawa, M., 2008. Decadal increase of anthropogenic CO₂ in the subtropical South Atlantic Ocean along 30°S. *J. Geophys. Res.* 113, C06007. doi:10.1029/2007JC004424.
- Orsi, A.H., Smeeth, J., W.M., Bullister, J.L., 2002. On the total input of Antarctic waters to the deep ocean: a preliminary estimate from chlorofluorocarbon measurements. *J. Geophys. Res.* 107, 3122. doi:10.1029/2001JC000976.
- Peng, T.-H., Wanninkhof, R., 2010. Increase of anthropogenic CO₂ in the Atlantic Ocean in the last two decades. *Deep-Sea Research* 57, 755–770. doi:10.1016/j.dsr.2010.03.008.
- Pérez, F.F., Vázquez-Rodríguez, M., Lounan, E., Padin, X.A., Mercier, H., Rios, A.F., 2008. Temporal variability of the anthropogenic CO₂ storage in the Irminger Sea. *Biogeosciences* 5, 1669–1679. doi:10.5194/bg-5-1669-2008.
- Pérez, F.F., Vázquez-Rodríguez, M., Mercier, H., Velo, A., Lherminier, P., Rios, A.F., 2010. Trends of anthropogenic CO₂ storage in North Atlantic water masses. *Biogeosciences* 7, 1789–1807. doi:10.5194/bg-7-1789-2010.
- Rios, A.F., Pérez, F.F., Fraga, F., 2001. Long-term (1977–1997) measurements of carbon dioxide in the Eastern North Atlantic: evaluation of anthropogenic input. *Deep Sea Res. Part II* 48, 2227–2239.
- Rios, A.F., Johnson, K.M., Álvarez-Salgado, X.A., et al., 2005. Carbon Dioxide, Hydrographic, and Chemical Data Obtained During The R/V Maurice Ewing Cruise in the South Atlantic Ocean (WOCE Section A17, 4 January–21 March 1994). *Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory. ORNL/CDIAC-148 NDP-084*, pp. 1–27.
- Rios, A.F., Vázquez-Rodríguez, M., Padin, X.A., Pérez, F.F., 2010. Anthropogenic carbon dioxide in the South Atlantic western basin. *J. Mar. Syst.* 83 (1–2), 38–44. doi:10.1016/j.jmarsys.2010.06.010.
- Sabine, C.L., Tanhua, T., 2010. Estimation of anthropogenic CO₂ inventories in the ocean. *Ann. Rev. Mar. Sci.* 2, 175–198. doi:10.1146/annurev-marine-120308-080947.
- Sabine, C.L., Key, R.M., Johnson, K.M., Millero, F.J., Poisson, A., Sarmiento, J.L., Wallace, D.W.R., Winn, C.D., 1999. Anthropogenic CO₂ inventory of the Indian Ocean. *Global Biogeochem. Cycles* 13, 179–198.
- Sabine, C.L., Feely, R.A., Gruber, N., Key, R.M., Lee, K., Bullister, J.L., Wanninkhof, R., Wong, C.S., Wallace, D.W.R., Tilbrook, B., Millero, F.J., Peng, T.-H., Kozyr, A., Ono, T., Rios, A.F., 2004. The oceanic sink for anthropogenic CO₂. *Science* 305, 367–371.
- Schlitzer, R., 2007. Assimilation of radiocarbon and chlorofluorocarbon data to constrain deep and bottom water transports in the world ocean. *J. Phys. Oceanogr.* 37, 259–276. doi:10.1175/JPO3011.1.
- Steinfeldt, R., Rhein, M., Bullister, J.L., Tanhua, T., 2009. Inventory changes in anthropogenic carbon from 1997–2003 in the Atlantic Ocean between 20°S and 65°N. *Global Biogeochem. Cycles* 23, GB3010. doi:10.1029/2008GB003311.
- Tanhua, T., Biastoch, A., Körtzinger, A., Lüger, H., Böning, C., Wallace, D.W.R., 2006. Changes of anthropogenic CO₂ and CFCs in the North Atlantic between 1981 and 2004. *Global Biogeochem. Cycles* 104 GB4017. doi:10.1029/2006GB002695.
- Tanhua, T., Körtzinger, A., Fris, K., Waugh, D.W., Wallace, D.W.R., 2007. An estimate of anthropogenic CO₂ inventory from decadal changes in ocean carbon content. *Proc. Natl. Acad. Sci.* 104, 3037–3042. doi:10.1073/pnas.0606574104.
- Tanhua, T., Olsen, A., Hoppema, M., Jutterström, S., Schirnack, C., Van Heuven, S., Velo, A., Lin, X., Kozyr, A., Alvarez, M., Bakker, D.C.E., Brown, P., Falck, E., Jeansson, E., Lo Monaco, C., Olafsson, J., Pérez, F.F., Pierrot, D., Rios, A.F., Sabine, C.L., Schuster, U., Steinfeldt, R., Stendardo, I., Anderson, L.G., Bates, N.R., Bellerby, R.G.J., Blindheim, J., Bullister, J.L., Gruber, N., Ishii, M., Johannessen, T., Jones, E.P., Köhler, J., Körtzinger, A., Metzl, N., Murata, A., Musielewicz, S., Omar, A.M., Olsson, K.A., de la Paz, M., Pfeil, B., Rey, F., Rhein, M., Skjelvan, I., Tilbrook, B., Wanninkhof, R., Mintrop, L., Wallace, D.W.R., Key, R.M., 2009. CARINA Data Synthesis Project. ORNL/CDIAC-157, NDP-091. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory. U.S. Department of Energy, Oak Ridge, Tennessee. doi:10.3334/CDIAC.org.ndp091.
- Tanhua, T., Steinfeldt, R., Key, R.M., Brown, P., Gruber, N., Wanninkhof, R., Pérez, F.F., Körtzinger, A., Velo, A., Schuster, U., Van Heuven, S., Bullister, J.L., Stendardo, I., Hoppema, M., Olsen, A., Kozyr, A., Pierrot, D., Schirnack, C., Wallace, D.W.R., 2010a. Atlantic Ocean CARINA data: overview and salinity adjustments. *Earth System Science Data* 2, 17–34. doi:10.5194/essd-2-17-2010.
- Tanhua, T., Van Heuven, S., Key, R.M., Velo, A., Olsen, A., Schirnack, C., 2010b. Quality control procedures and methods of the CARINA database. *Earth System Science Data* 2, 35–49. doi:10.5194/essd-2-35-2010.
- Thomas, H., Ittekkot, V., 2001. Determination of anthropogenic CO₂ in the North Atlantic Ocean using water mass ages and CO₂ equilibrium chemistry. *J. Mar. Syst.* 27, 325–336.
- Touratier, F., Azouzi, L., Goyet, C., 2007. CFC-11, Δ14C and 3H tracers as a means to assess anthropogenic CO₂ concentrations in the ocean. *Tellus* 59B, 318–325. doi:10.1111/j.1600-0889.2006.00247.x.
- Van Heuven, S.M.A.C., Hoppema, M., Huhn, O., Slagter, H.A., de Baar, H.J.W., 2011. Direct observation of increasing CO₂ in the Weddell Gyre along the Prime Meridian during 1973–2008. *Deep-Sea Research* 58, 2613–2635.
- Vázquez-Rodríguez, M., Padin, X.A., Rios, A.F., Bellerby, R.G.J., Pérez, F.F., 2009a. An upgraded carbon-based method to estimate the anthropogenic fraction of dissolved CO₂ in the Atlantic Ocean. *Biogeosciences Discuss.* 6, 4527–4571. doi:10.5194/bgd-6-4527-2009 (<http://www.biogeosciences-discuss.net/6/4527/2009/bgd-6-4527-2009-print.pdf>).
- Vázquez-Rodríguez, M., Touratier, F., Lo Monaco, C., Waugh, D.W., Padin, X.A., Bellerby, R.G.J., Goyet, C., Metzl, N., Rios, A.F., Pérez, F.F., 2009b. Anthropogenic carbon distributions in the Atlantic Ocean: data-based estimates from the Arctic to the Antarctic. *Biogeosciences* 6, 439–451. doi:10.5194/bg-6-439-2009.
- Velo, A., Vázquez-Rodríguez, M., Padin, X.A., Glicoto, M., Rios, A.F., Pérez, F.F., 2010. A multiparametric method of interpolation using WOA05 applied to anthropogenic CO₂ in the Atlantic. *Sci. Mar.* 74 (S1), 21–32. doi:10.3989/scimar.2010.74s1021.
- Wanninkhof, R., Doney, S.C., Bullister, J.L., Levine, N.M., Warner, M., Gruber, N., 2010. Detecting anthropogenic CO₂ changes in the interior Atlantic Ocean between 1989 and 2005. *J. Geophys. Res.* 115, C1028. doi:10.1029/2010JC006251.
- Waugh, D.W., Hall, T.M., McNeil, B.L., Key, R.M., Matear, R.J., 2006. Anthropogenic CO₂ in the oceans estimated using transit time distributions. *Tellus* 58B, 376–389. doi:10.1111/j.1600-0889.2006.00222.x.